UROP Undergraduate Thesis

Identifying Fungal Pathogens in the Air of Atlanta

Konstantinidis Laboratory
(www.enve-omics.gatech.edu)

School of Civil and Environmental Engineering

Casey Erb

Spring 2020

Faculty Member 1

_____

Faculty Member 2

_____

Table of Contents

Abstract

The atmosphere is a vastly understudied habitat for airborne microbial communities invisible to the naked eye. Very little is known regarding the microbial composition of these airborne communities and how composition varies across different meteorological conditions. Even less is known regarding the potential impact of these bioaerosols on human health. Capturing a representative sample of the microbes present in the air is technically challenging, and traditional culture-based methods often capture <1% of the total airborne cells. To circumvent these limitations of culturing, this study employed nucleic acid-based analysis of microbial cells recovered directly from the sample filter. The goal of the study was to provide snapshots of the airborne microbial community in Atlanta's air throughout a time series of two consecutive years. Polymerized chain reaction (PCR) was used to amplify the highly conserved ITS regions of fungal genomes. The resulting sequences were analyzed using established bioinformatics pipelines in order to identify the microbes present. The analysis revealed several fungal species representing common pathogens of plants to be present in these samples as well as species associated with respiratory events in humans with asthma or other upper respiratory conditions. Future work to expand the time series and the breadth of the study to include viruses could answer the epidemiological mystery surrounding the cause of seasonality in respiratory infections (e.g., whether or not is airborne), and it could build a more complete picture of the inherent health risks of breathing open air.

Acknowledgements

Thank you to the Enve-omics lab for giving me the opportunity to do research. Thank you Kostas Konstantinidis, Lizbeth Davila, Janet Hatt, Blake Lindner, Arnaldo Negron, and Jacob Jackson for all of your help. This project would not have been possible without their support, and they will continue to develop this field of work.

Introduction

Bioaerosols are airborne microorganisms (i.e. bacteria, fungi, and viruses) that interact with each other and the surrounding environment. Airborne microorganisms rarely exist as singular particles in the air. Instead, they tend to aggregate with each other depending on individual size, thermodynamic conditions, and concentration [1]. Particulate matter also aggregates with bioaerosols in the atmosphere and can influence the infectivity of pathogenic bioaerosols [2]. The behavior, quantity, and identity of bioaerosols play a significant role in the transmission of disease, but microorganisms in ambient air tend to exist at concentrations that are orders of magnitude lower than in other habitats [3]. As a result, collecting bioaerosol samples with a significant amount of biomass for analysis is a technical challenge. Bioaerosol particles also have low inertial mass that is challenging to collect in a sampling medium [1]. Epidemiological researchers have led the way in developing technology to study airborne pathogens and other bioaerosols [1].

Extensive research has been conducted on the quantity and identity of bioaerosols indoors, with a focus on pathogenic organisms. The infectivity and airborne concentration of viruses such as Influenza and SARS coronavirus have been observed in healthcare settings (e.g. hospitals) due to public concerns surrounding nosocomial infections [4] [5] [6]. Nosocomial infections are those that occur between patients in a hospital due to improper sanitation. Sampling pathogenic bioaerosols in confined spaces broadens our understanding of the transmission of a disease, and the relatively high concentration of airborne pathogens in hospital settings makes collecting adequate biomass less complex [4]. These previous studies have found that although routine sanitization is a necessary precaution, it does not guarantee an absence of pathogenic bioaerosols [4]. In outdoor settings, research has been performed to study the behavior of bioaerosols in the atmosphere (settling velocity, movement, and travel) and estimates ambient concentrations of bacterial and fungal biomass [3].

However, the current state of research in this field fails to describe the quantity and identity of pathogenic bioaerosols in the open air. Even less is known regarding the influence of seasonal and meteorological forces on the atmospheric concentration of pathogenic bioaerosols despite the consistent epidemiological trends of seasonal infection for most pathogens [7]. For instance, no scientific consensus exists as to whether or not the seasonal infection rates of pathogens such as Influenza, Rubella, and Rotovirus are caused by an increased ambient air concentration of the pathogens [7]. Alternative explanations of the phenomena include seasonal changes in host population density, pathogen virulence, and host immune response due to weather conditions [7]. No matter the true cause of seasonal infectivity, a strong baseline of pathogenic quantity and identity could help inform the methods public health officials use to reduce seasonal infections.

This study sampled ambient air and rainwater in the Atlanta metro area, and used microbial tools to quantify and identify pathogenic microorganisms. The samples establish a time series of two years, and results were analyzed for seasonal fluctuations and the impact of environmental conditions. Traditional culturing methods are known to be inefficient for

analyzing airborne organisms, so established techniques for nucleic acid-based research were implemented instead. Due to weak biomass signals in atmospheric samples, sterility and careful handling were of paramount importance throughout the sampling, extracting, amplifying, purifying, and sequencing stages of nucleic acid analysis. This study amplified and sequenced the "barcode" Internal Transcribed Spacer (ITS) region of fungal genomes in order to identify the fungal microbes present in each sample. In densely populated urban areas like Atlanta, the impact of this field of research on public health practices is particularly significant and could help towards preventing the spread of infections and saving lives. Establishing a clear estimate of the proportion of pathogenic bioaerosols is the first step to understanding the risk of infection from interacting with the open air.

Literature Review

Respiratory infections can range in severity from the common cold to tuberculosis, and there are over one billion cases of respiratory infection each year in the U.S. alone [9]. The symptoms and direct transmission of these infectious diseases are well understood, and preventative practices to avoid contact with pathogenic microorganisms are commonplace. Pathogens can spread through direct or indirect contact, ingestion, or inhalation in aerosolized form [9]. Pathogenic bioaerosols are formed when infected hosts cough or sneeze, suspending microbes in the air. The size of these pathogenic bioaerosols varies from as small as 0.02 μm in diameter for some viruses to as large as 5 μm in diameter for some bacteria and fungi, and anything larger is classified as an airborne droplet [6]. Plant pollen is an exception to this rule and is classified as a bioaerosol despite being larger than 5 μm in diameter.

Because the settling velocity of an airborne particle is proportional to its surface area, small particles can stay suspended in the air for surprisingly large amounts of time [3]. For example, a viral particle 0.03 μm across aerosolized 2 m off the ground takes over 3 months to settle [3]. Consequently, every breath of air contains an unknown quantity of microbial biomass and an unknown proportion of pathogens. Current estimates for total bacterial and fungal concentration in open air vary from 10^2 to 10^6 colony forming units per m^3 for bacteria, and 10^2 to 10^3 spores per m^3 for fungi [3]. However, due to the technical challenges of sampling and analyzing microbes in this habitat, no consensus exists as to the concentration and identity of pathogenic microbes in the open air. Studies directly focused on answering this question often opt for a smaller scope, focusing on a specific pathogen in a controlled environment. The setting of the sample, the sampling method, and the approach to nucleic acid analysis all influence the accuracy of the final estimation of pathogenic quantity. The current study used a combination of validated methods by previous studies in order to establish a time series of informative snapshots of the airborne microbial community. Ultimately, the potential for further research in this field could have profound impacts on the methods humans use to combat and prevent infections.

Indoor Studies

High-risk indoor environments such as hospitals have historically been prioritized settings for studies that sample bioaerosols. A study by Booth et al. has tested several sampling methods in hospital settings [4]. The study examined the presence of Severe Acute Respiratory Syndrome (SARS) coronavirus in patient rooms using both dry air and wet air sampling (along with surface swabs) and found few positive samples [4]. The purpose of this study was to identify whether the rigorous sanitation practices employed by the hospital were effective at sterilizing patient rooms of SARS particles. Dry air sampling involved pumping approximately 1,400 L of air through polytetrafluoroethylene (PTFE) filters and wet air sampling involved slit centrifugation and collection in a liquid medium [4]. Both sampling methods yielded similar negative results and very few SARS particles were captured. The study did not provide an estimate for the concentration of SARS particles in patient rooms that were not sterilized or in rooms that did not have patients in them [4].

Recently, in order to test the efficiency of emerging technologies in collecting pathogenic bioaerosols, studies have purposefully introduced a pathogen to an unsterilized indoor environment. A study focusing on foot-and-mouth disease (FMD) aerosolization from piglets was conducted in an enclosed barn environment by Medaglia et al. [10]. Foot-and-mouth disease represents one of the smallest viral particles at 0.02 μm [10]. As previously mentioned, it is theoretically difficult to force extremely small particles into a sampling medium because of their low inertia and susceptibility to being buffeted erratically by gaseous particles [1]. Medaglia et al. obtained positive identifications of FMD using a slit cyclone sampler, and they quantified the number of FMD molecules in each sample using quantitative PCR. These authors found PCR gene copies to vary unpredictably with sampling time, suggesting that even in a controlled indoor environment, microbial biomass is difficult to quantify [10]. 2-hour samples were found to contain between 2,160-3514 FMD particles [10].

It should be noted that this quantity of particles is adequate for identifying a known pathogen, but if nucleic acid sequencing is to be performed on multiple unknown pathogens, the recommended minimum number of gene copies recovered is roughly 10^{11}[3]. This estimate accounts for losses in purity due to extraction inefficiencies and the natural impurities present in aerosol samples [3]. The works of Medaglia et al. and Booth et al. suggest that a slit cyclone sampling instrument is capable of capturing even the smallest bioaerosols. The study suggests that this instrument may be viable for collecting certain pathogens in samples of outdoor environments. However, the small sample size and low particle counts suggest that this method may produce highly variable results [10].

Amplification and Analysis: quantitative PCR (qPCR)

Once bioaerosols have been collected, amplification is performed in order to study their nucleic acids (molecular identification and quantification). A common and reliable (albeit somewhat complicated) method of analyzing nucleic acid is through a quantitative polymerized chain reaction (qPCR) assay [11]. The primary objective of qPCR analysis of airborne samples is to arrive at a volumetric concentration of biomass denoted as C_t [11]. Scientists working for

Applied Biosystems (a manufacturer of qPCR instruments) have performed research to determine what "template-independent factors" can influence the calculation of C_t when using qPCR. The study described potential loss of nucleic acid during any manipulation of the nucleic acid before the qPCR assay is run [11]. Viral nucleic acid in particular requires more manipulation than bacterial or fungal nucleic acid prior to conducting PCR. If a viral species uses RNA as its primary genetic code, then the extracted nucleic acid must be reverse transcribed prior to conducting PCR of any kind [12]. Viruses also do not have a highly conserved genomic region like bacteria's 16S rRNA gene or fungi's ITS (internal transcribed spacer- region between the 18S and the 25S rRNA genes), so qPCR assays for viruses must be conducted with primers and probes that target a single species or group of related viruses. Accordingly, the specificity of these primers and probes (e.g., existence of nucleotide mismatches) can also affect the final C_t [11].

After primers and probes are selected, double stranded DNA is ready for amplification. The qPCR process past this stage is identical for bacteria, fungi, and viruses. A fluorescent probe containing an annealing sequence specific to the desired amplicon is used to amplify the (target) nucleic acid. A qPCR instrument can then detect the strength of the fluorescent signal and produce an estimate of sample biomass concentration (or copy number), C_t. As with any spectrophotometry experiment, a standard curve is created using a series of samples with previously known fluorescence (or copy numbers). In this case, the fluorescence is directly proportional to a known copy number of the target DNA [11]. The standard curve is also used to establish the limit of detection: the weakest signal that can be detected by the instrument. Researchers have found that the efficiency of the PCR, the precision of the outputs, and a lack of sensitivity (the inability of a single cell's signal to amplify significantly) can all lead to inaccuracies when observing C_t [11]. A significant challenge in studying the ITS region of the fungal genome arises when using qPCR because fungi may have multiple instances of the ITS gene. This can lead to inaccurate cell counts, or at best, semi-quantitative results.

Advantages of using qPCR with Bioaerosols

This method certainly has its merits when it comes to diagnosing patients in real time. A study by van Elden et al. has confirmed that qPCR is more specific and gives less false positives than conventional diagnostic viral culturing methods [13]. qPCR is extremely effective for this application because it requires no post-PCR analysis to confirm the presence of the pathogen and can give results within 4 to 5 hours from the sampling time [13]. A study by Fabian et al. has optimized the bench work process of detecting a positive signal for Influenza and human rhinovirus using Trizol-chloroform extraction, a specific brand of reverse transcriptase kit, and qPCR [14]. To study potential inhibition of nucleic acid amplification by inorganic matter, samples with known concentrations of Influenza and rhinovirus were deliberately spiked with up to 50 µg of typical atmospheric particulate matter (dust and clay particles) [14]. No inhibitory effects were observed below the 50 µg threshold [14]. Fabian et al. study provides a promising post-collection protocol for the successful identification of

specific viruses in samples of ambient air with low bioaerosol concentration and high particulate matter concentration.

Limitations of using qPCR with Bioaerosols

It seems that when seeking to confirm the presence of a single, known pathogen, qPCR is the optimal method, especially when working with airborne viruses as mentioned above. qPCR falls short, however, of capturing a "snapshot" of the entire microbial community because it requires a probe to be prepared for each species of interest. The World Health Organization (WHO) suggests to use qualitative PCR to amplify nucleic acid, using the same primers as qPCR without probes, then using gel electrophoresis to confirm the presence of amplification [12]. PCR represents a more qualitative approach to identify a viable signal of biomass for sequencing.

Gel electrophoresis has many advantages as an efficient and cost-effective method to confirm the presence of a signal and identify the length of amplicons from PCR. Only a small amount of PCR product is needed to run a gel electrophoresis, and if a signal is confirmed, the rest of the PCR product can be purified and sequenced [12]. The WHO protocol was originally developed to confirm the presence of Influenza by comparing the known length of the region with the amplicon's length [12]. PCR method is not as specific as qPCR for identification of single virus species (it will generate more false positives). The fluorescent dye used in gel electrophoresis is added after PCR and will bind to any amplicon, whereas the fluorescent probes used in qPCR bind only to amplicons of the virus species. Qualitative PCR has utility when identifying the existence of a signal in samples of multiple unknown bacterial or fungal species.

Fungal ITS Primers

A common pitfall when performing PCR of any kind is introducing error by using incorrect primers. A study by Ihrmark et al. used gel electrophoresis to analyze 11 mock (i.e., of known composition) samples of common fungal microbes [15]. The study tested ITS primers fITS7, gITS7, and fITS9 to determine their effectiveness relative to the more commonly used ITS1f/ITS4 primer combination [15]. The lengths of the amplicons were compared to the known length of the ITS region for each of the 11 species, and it was found that the traditional ITS1f/ITS4 combination had a bias against species with longer amplicons [15]. The study also used gel electrophoresis on natural wood, wheat, soil, and hay samples using the new combination of primers fITS7, gITS7, and fITS9 [15]. The samples were eventually sequenced to identify the fungal species present, and the use of gel electrophoresis to confirm a signal constituted a more cost-effective approach than qPCR.

No precise quantitative measure of concentration can be obtained by observing the gel, but the study used an alternative instrument, a Qubit Fluorometer, to determine the concentration of DNA in the PCR product. This instrument uses a light-sensitive, fluorescent dye that is added to the PCR product after amplification, rather than expensive, species-specific probes that are added before amplification in the case of qPCR [15]. Using qualitative PCR, gel electrophoresis, and the Qubit Fluorometer allows for the length of

amplicons and the concentration of nucleic acid to be observed in a cost- effective manner that detects all species present in the sample. This approach is more cost effective (but less accurate, in general) than qPCR for identifying multiple bacterial and fungal species. Unfortunately, the study provided no evidence regarding the effectiveness of this new set of primers on nucleic acid from samples of the atmosphere. Airborne microbial communities differ in size and composition from those in soil and vegetation, so it is possible that the efficacy of this new set of primers (in terms of capturing the diversity of species present in the sample) would not translate to atmospheric samples. In addition, there might be (different) inhibitors of PCR in air samples due to the particles and organics sampled as parts of bioaerosols.

Summary Points from Established Literature for the Present Study

This study analyzed a time series of air and rainwater samples to observe trends in the quantity and identity of microorganisms (both pathogenic and benign) in the atmosphere. The study utilized the nucleic acid-focused analysis techniques established by previous studies to characterize airborne microbial communities. Specifically, it employed three sampling methods, one of which is the slit cyclone sampler described in the work of Medaglia et al. Further, the study used the primers described in the study by Ihrmark et al. for fungal analysis, and followed the WHO protocol for qualitative PCR. It also used a Qubit fluorometer rather than qPCR to obtain semi-quantitative readings, as implemented by Ihrmark et al [15]. The novelty of the study is primary due to the unique samples obtained, which represent a two year time series of the air of the metro area of a major Southeastern city. Such time series do not exist to the best of our knowledge at the time of this writing. The study serves as a foundation for characterizing the risk of adverse health effects from breathing open air.

Materials and Methods

Summary

Samples of airborne bacteria and fungi were taken using polytetrafluoroethylene (PTFE) filters and high volume sampling on the roof of the Environmental Science and Technology building on Georgia Tech campus. Rainwater was also collected in previously cleaned and autoclaved jars. Nucleic acid was extracted from samples the highly conserved 16S ribosomal ribonucleic acid (rRNA) region of bacterial genomes or the highly conserved ITS region of fungal genomes were amplified via PCR. Purifying and sequencing the PCR product identified bacterial and fungal microbes (including pathogens) in samples.

Protocol

Two methods were used for sampling dry air: a High Volume (Hi-Vol) sampler and a slit cyclone sampler. Both instruments were set up on the roof of the Ford Environmental Science and Technology Building on Georgia Tech’s campus. Figure 13 in the Appendix shows the Hi-Vol sampler, manufactured by Fisher Scientific.

Dry Air Collection Method 1

The Hi-Vol sampler is a vacuum-powered instrument designed by Thermo-Scientific to filter ambient air. It uses a pump to pull air into the inner chamber through 10 µm inverted slits. The inner chamber contains a mesh screen that holds an initially sterile PTFE filter in place. For all Hi-Vol samples used in this study, the pore size of the PTFE filter was 0.8 µm. This is not strictly small enough to capture smaller viral bioaerosols, but it captures fungi and bacteria efficiently. The Hi-Vol's interior was cleaned thoroughly with 70% ethanol prior to each collection. Hi-Vol collections were run for 24 hours during days with no precipitation, since water pulled into the inner chamber could damage the pump. The pump filters air at 40 ft^3/min, leading to a total of 1,630,000 L of air filtered throughout the 24-hour window. After collection, the filter was cut into strips and suspended in sterile Phosphate-Buffered Saline (PBS) solution throughout a pre-treatment protocol (see Table 9 in Appendix for PBS components). This protocol was designed to physically agitate the strips of filter using a VWR Pulsing Vortex Mixer, so that the particulate matter captured became suspended in the PBS buffer. This buffer was then filtered using a bench vacuum pump and a smaller Millipore filter with a pore size of 0.2 µm.

Dry Air Collection Method 2

The slit cyclone sampler (SpinCon) is manufactured by InnovaPrep and captures aerosols in a liquid medium of (initially sterile) PBS buffer. The sampler pulls air through a tube and traps particles that travel through a slit in a centrifugal chamber. The liquid medium swirls continuously through the chamber, while the air passes out the exhaust pipe. The SpinCon can sample at differing flowrates, but was set to 450 L/min for these samples. The samples in this experiment all had a collection time of 4-6 hours, meaning a total of 108,000-162,000 L of air were filtered. The Spincon is designed to give a final elute volume of approximately 40 mL. No pre-treatment was necessary for samples collected by the SpinCon sampler, and they were immediately filtered using the bench vacuum pump onto a filter with a pore size of 0.2 µm. The SpinCon is a delicate, precise instrument that must be cleaned very carefully. After all accessible surfaces were cleaned with 70% ethanol, there was an iterative cleaning process that runs bleach, ethanol, water, and PBS buffer through the internal tubing of the instrument to ensure there is no contamination [18]. This was performed immediately before sampling.

Rainwater Collection

Rain samples were also collected from the roof of the Ford ES&T building. Three autoclaved (sterilized) 1 L glass jars were set up with funnels to collect rain during periods of heavy precipitation. The funnels were cleaned thoroughly with Alconox (a detergent) and wiped down with 70% ethanol immediately prior to the collection period. Rain collection time periods varied with the intensity of the rain. Samples with at least one total Liter of rainwater were used for analysis. Some collections fill all three jars completely while others are closer to the minimum mark. The total volume collected during each event was recorded. Rainwater was filtered directly onto larger Whatman filters with a 0.2 µm pore size, and these filters were used for DNA extraction.

Sample Metadata

We collected the same metadata for each sample. Specifically, the time and date at the beginning and end of each sampling period were noted, and notes were maintained for any deviations from normal conditions. Weather conditions are monitored before, during, and after sampling, and National Oceanic and Atmospheric Administration (NOAA) records were used to create consistent records of meteorological factors that could impact the quantity and identity of the microorganisms observed. When the nucleic acid of samples was not immediately extracted, the samples were stored in petri dishes and sterile bags in a -80˚C freezer.

Extraction

The nucleic acid extraction protocol was a three day long, phenol chloroform extraction. The process was derived from a protocol developed for a previous study by the Enve-omics lab [19]. It was a deliberately manual extraction protocol so that the purity of the samples was preserved and no contamination occurred, which is common with commercial DNA/RNA extraction kits due to impurities. A summary of the extraction protocol is given in the Appendix.

PCR

Once nucleic acid was extracted from samples, the 16S rRNA region of bacterial genomes and the ITS region of fungal genomes was amplified via PCR. The protocol of the PCR is included in the Appendix for replicability. A negative control was included during each batch of PCR to ensure no contamination occurred, and batches with amplification on the negative control were not included in the final results. Each negative control contained the primers and other reagents added before and after amplification, but had an extra 1.25 µL of Millipore water instead of the 1.25 µL of sample DNA template. All reagents, working surfaces, and containers were either autoclaved, sterilized with 70% ethanol, or UV treated to eliminate all possible sources of false positives. In the Appendix, Table 10 shows the primers used for fungal ITS amplification, Table 11 shows the PCR mix components in their respective proportions, and Table 12 shows the temperatures and runtimes in the thermocycler.

Gel Electrophoresis

The quality of PCR runs was checked using gel electrophoresis. Dye was loaded with isolates of the PCR product, and the length of the amplicons was observed using a UV light. Gel electrophoresis also provided insight into whether the negative control amplified during PCR. If a band appeared in the gel for the negative control, then it is possible that the PCR mix was contaminated before amplification. Only PCR products with no fluorescence on the negative controls were sequenced in this experiment. No negative controls were sequenced as blanks due to outputs of a BioAnalyzer run, which indicated that the nucleic acid presence in multiple negative controls was entirely due to primer dimer.

Purification

The PCR products were purified (the amplicons of the correct length were isolated) using the SPRI Select magnetic bead purification kit. This removed amplicons that were a result of the primers finding themselves and amplifying. The protocol for SPRI Select bead purification was

followed exactly from the SPRIselect User Guide developed by Beckman Coulter. The result of the size selection step of the protocol was a concentration of beads that was 0.8x the concentration of PCR product. A Qubit fluorometer was used to quantify the amplified DNA before the samples were sent for sequencing.

Sequencing and Community Analysis

Sequencing of amplicon was performed on an Illumina MiSeq instrument running the 250 paired-read kit. Qiime 2 version 2018.8 was used to analyze the reads. First, forward and reverse reads were merged into one sequence when overlapping, using Qiime 2's import function and Casava One Eight Single Lane format. Qiime 2's vsearch function was then used to de-replicate reads. The sequences were then matched to organizational taxonomic units (OTUs) using vsearch's cluster feature at 99% identity. Each OTU was a representative sequence from the NCBI database. Qiime 2's classifier feature and the sklearn function were then used to create complete taxonomic classifications for each OTU (from kingdom down to the most specific level possible). The classifier is a machine learning algorithm that operates based on known pairings of taxonomic identifications and representative sequences of base pairs (for fungal analysis, the UNITE database was used for reference). The classifier used was gg-13-8-99-515-806-nb-classifier.qza, which is a part of the Qiime 2 package. Qiime 2 then allows for biodiversity analysis among the samples as well as against the collected metadata using the diversity core metrics feature. The sampling depth used for diversity analysis was 9,600 sequences per sample, based on the alpha rarefaction curve (Figure 14 in the Appendix) and the sample with the lowest number of reads. Together, the time series of samples can illustrate how the bioaerosol community changes with seasons and meteorological conditions.

Blastn Analysis

After using the Qiime2 pipeline to study the general trends of the fungal bioaerosol community in our samples, a more detailed analysis was conducted using Blastn to identify a pathogenic or allergenic fraction of the community. A database of 103 putative fungal pathogens and allergens was assembled through review of literature [16,17]. The original reads were re-paired using a software called PEAR. PEAR creates merged .fasta files for each sample, which are the input to the Blastn feature. Blastn was used to identify the relative abundance of each of these 103 pathogens in each sample (threshold for a match: 97% identity and alignment length >200 basepairs). These thresholds were chosen to give a more confident indication of species-level matching, and 23 species were found to be present in at least one sample. Next, a distance matrix was created only for this small matching pathogenic fraction, and Adonis analysis was repeated on this reduced dataset. The objective of this second Adonis analysis was to narrow the focus and identify the environmental factors that influenced the presence or absence of the 23 putative pathogens identified in the samples.

Results

The raw sequencing reads were first merged, and quality trimmed. Figure 1 shows a quality plot of fungal ITS sequences from 60 samples, indicating the accuracy of the base calls

along the length of the average read. A sharp decrease in quality calls was noticed but it was within the specifications of the Illumina sequencer instrument for a good run and the overall read quality was good.

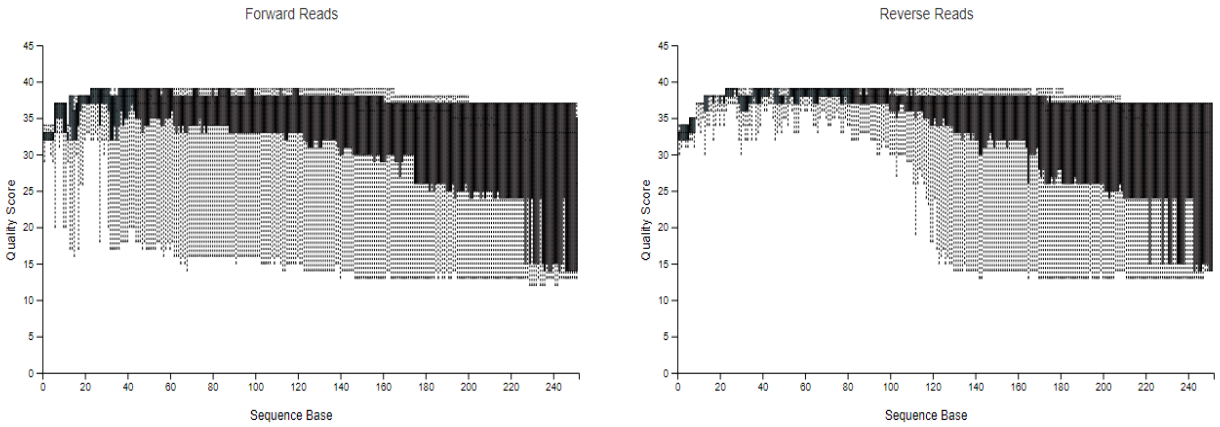


Figure 1: Sequence Read Quality Plot Generated in Qiime 2.

The data was then assigned into organizational taxonomic units, or OTU's, a proxy for species. Occasionally, individual OTU's cannot be assigned to species or genus level, because different species/genera share an identical ITS gene sequence at this level or the species the sequences represent is an unknown (not previously characterized) taxon. The NCBI database was also used to link OTU to specific taxa.

Figure 2 illustrates the relative abundance of the taxa among the read sequences of the 54 samples analyzed in total. The full taxonomic name of each observed species (or genus, etc) is found using a machine learning-based classifier within Qiime 2, which references the UNITE fungal database rather than the NCBI database in this case. Figure 2 shows the genus-level classifications; Figure 3 shows the species level when available.

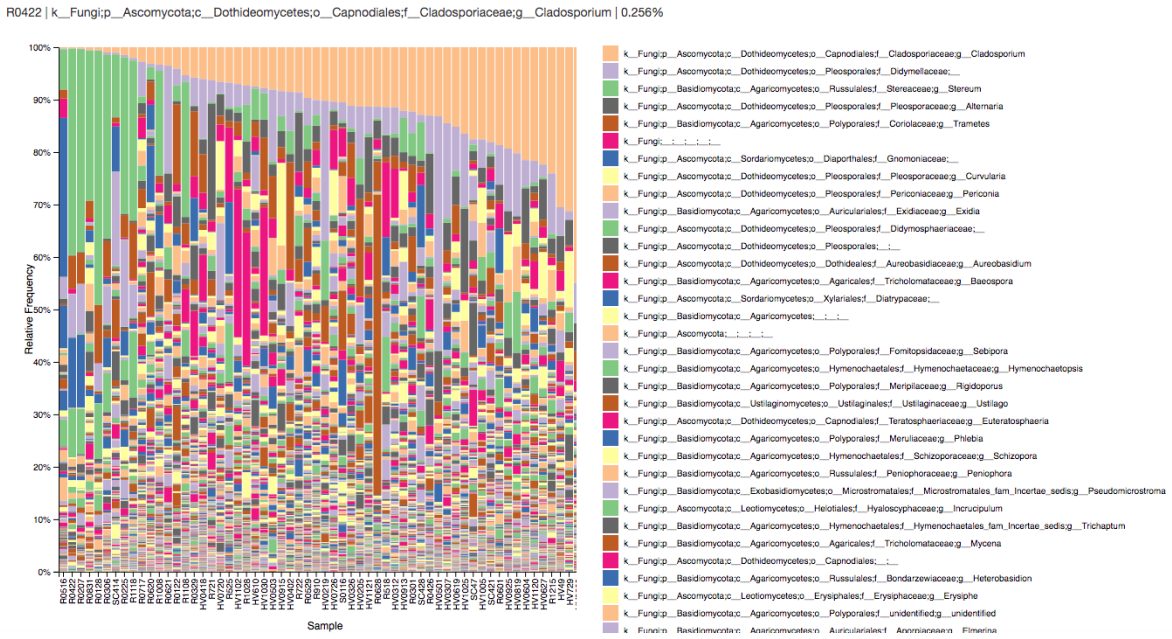


Figure 2: Genus-level Barplot Generated in Qiime 2

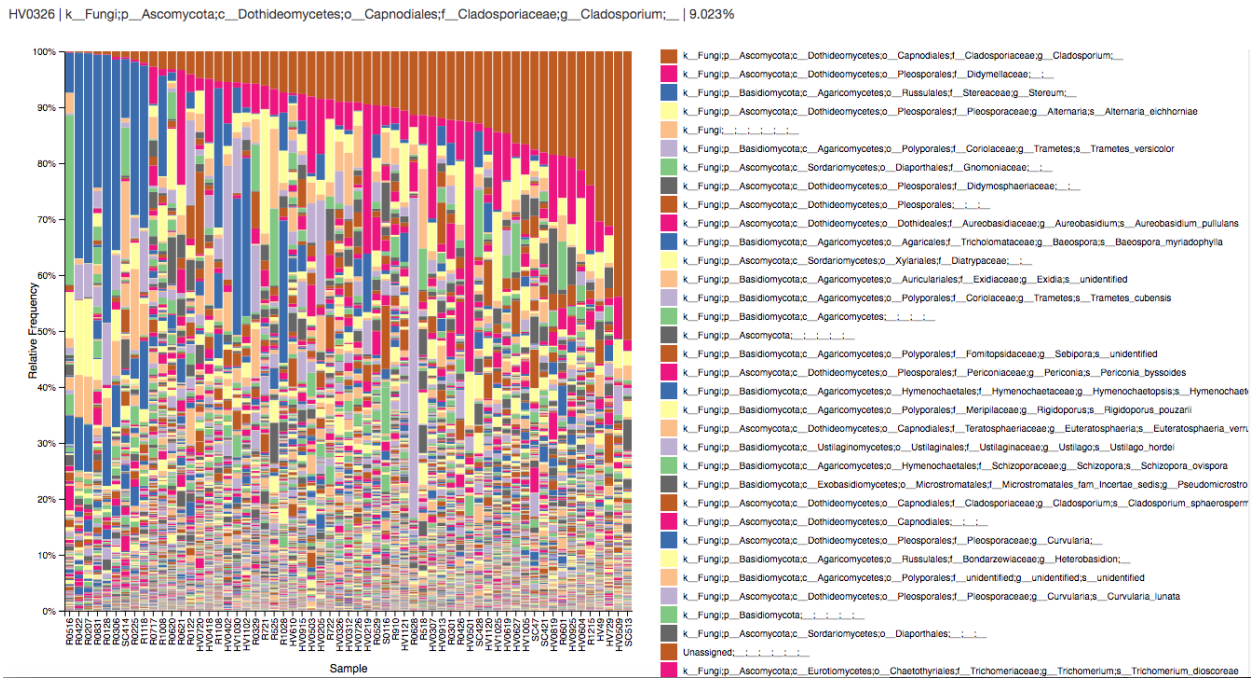


Figure 3: Species-level Barplot Generated in Qiime 2

The relative abundance data was used to create a Bray Curtis distance matrix to study the beta-diversity patterns among samples. This matrix demonstrates how correlated the fungal communities of each sample are with every other sample, using values that represent normalized phylogenetic distance.

Table 1 shows the Adonis analysis of the Bray-Curtis distance matrix. The Adonis analysis computes the R^2 value (correlation coefficient) between any external variable and the values in the Bray-Curtis distance matrix. The R^2 column represents the percentage of variation in normalized inter-sample biodiversity "distance" that can be statistically explained by a certain variable. It is clear from the results of the Adonis analysis that none of the environmental factors alone can explain a major fraction of the changes in the fungal bioaerosol community. Relative abundance of taxonomic groups seemed to vary the most by the origin of the airmass (back trajectory), along with the season and sampling method used (Hi-Vol, slit cyclone, or rain). Each of these factors explains about 10% of the variation among the samples. The highest R^2 value for an individual influence was airmass origin, with a value of 0.116. This means that approximately 11.6% of the variation in phylogenetic distance between samples can be explained by the cardinal origin of the airmass. Collectively, these results suggested that fungal bioaerosol communities vary more significantly with geography and time than with day-to-day weather changes. The small R^2 values of the quantitative variables (temperature, rainfall, pressure, humidity, PM counts, ozone, and pollen counts) indicate that, in isolation, these factors are poor predictors of the bioaerosol community. Each of these quantitative factors alone can only explain 1-2% of the phylogenetic distance between samples. Overall, our results suggested that 44.1% of the variability among fungal communities sampled could be explained by the 12 variables

individually, and 100% of the variability was explained by linear combinations of the 12 parameters.

Table 1: Adonis Analysis of Fungal Bioaerosol Community

	Df	SumsOfSqs	MeanSqs	F.Model	R2	Pr(>F)
Season	3	1.5926	1	0	0.09451	1
SamplingMethod	3	1.2912	0	0	0.07663	1
BackTrajectory	7	1.9536	0	0	0.11593	1
Pressure	1	0.3100	0	0	0.01839	1
Temperature	1	0.2710	0	0	0.01608	1
Rainfall	1	0.2656	0	0	0.01576	1
RelHumidity	1	0.2861	0	0	0.01698	1
SolarFlux	1	0.2263	0	0	0.01343	1
PM2Point5	1	0.2571	0	0	0.01526	1
PM10	1	0.3651	0	0	0.02167	1
Ozone	1	0.2120	0	0	0.01258	1
Pollen	1	0.3981	0	0	0.02363	1
Season:SamplingMethod	3	0.6162	0	0	0.03657	1
Season:BackTrajectory	11	2.9009	0	0	0.17216	1
SamplingMethod:BackTrajectory	8	2.6252	0	0	0.15579	1
Season:Pressure	3	0.8445	0	0	0.05011	1
SamplingMethod:Pressure	2	0.6726	0	0	0.03992	1
BackTrajectory:Pressure	5	1.2986	0	0	0.07707	1
Season:Temperature	1	0.4639	0	0	0.02753	1
Residuals	0	0.0000	Inf		0.00000	
Total	55	16.8507			1.00000	

The NOAA Hysplit model was used to collect metadata describing the ambient conditions on the day of each sample, and the origin of the airmass collected (oceanic vs. land transport). The Hysplit model generates a back-trajectory of an air mass for the three days prior to each sample, and tracks variables including the cardinal origin of the air mass, temperature, rainfall, relative humidity, and solar intensity, at half hour intervals. These metadata values were then used in the ANOVA of dissimilarities (ADONIS) analysis in RStudio to correlate OTU differences between samples to environmental variables. The airmass origin included in the metadata is based on the visual inspection of the back-trajectory plot generated by the Hysplit model. The quantitative variables in the metadata were average values for the previous 3 days.

Figure 4 shows an example output plot of the NOAA Hysplit Model for sample R0122, a rain sample (denoted by “R” in the sample id) taken on January 22nd, 2018 (01 for the month and 22 for the date in the sample id). The output plot was used to gather metadata for the Adonis analysis. The red, blue, and yellow lines indicate the back-trajectory values at 50, 100, and 500 meters above ground level, respectively. The cardinal direction used as the back-trajectory for the metadata (N, NE, E, etc.) was based on qualitative observation of these three lines. For example, for sample R0122, the back-trajectory assigned was “South.” The quantitative metadata from the NOAA Hysplit model represented the average of the 50 m AGL data over the previous 3 days.

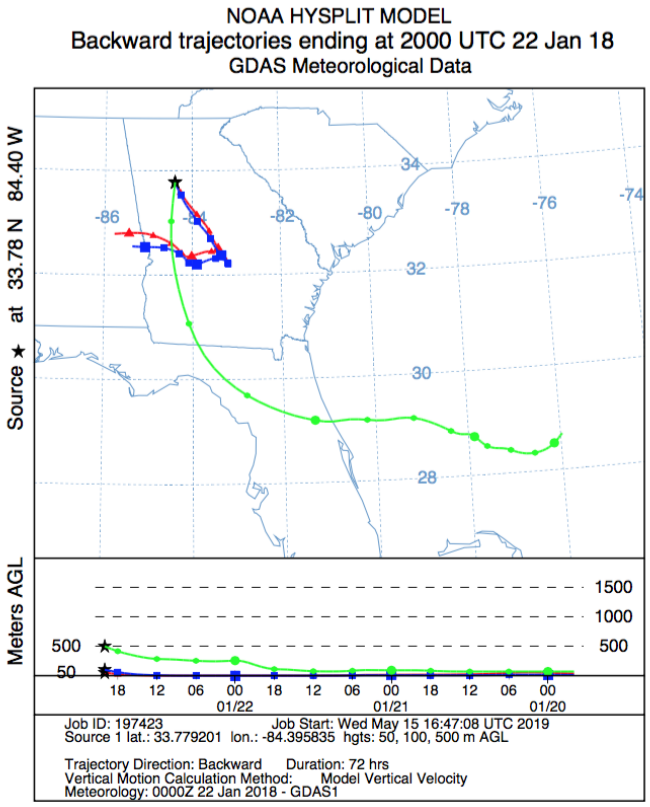


Figure 4: NOAA Hysplit Plot for Jan 22, 2019 Sample showing the back trajectory of the air masses sampled at different altitudes above ground level (represented by the three color lines). See text for further details.

Table 2 shows the collected metadata used for the Adonis analysis. Additional environmental variables to those available as part of the NOAA Hysplit report were collected for each sample day. The total 12 variables available were correlated with the distance matrix output from Qiime2. The variables included: Season, Sampling Method (Rain, Hi-Vol, or Spincon), Back-Trajectory (N, NE, E, etc.), Pressure (mm hg), Temperature (K), Rainfall (mm per hour), Relative Humidity (%), Solar Radiation Flux (W/m^2), $PM_{2.5}$ ($\mu g/m^{3)}$, PM_{10} ($\mu g/m^3$), Ozone (ppb), and Pollen Count (g/m^3).

Table 2: Metadata for Adonis Analysis

Sample-ID	input	filtered	denoised	non-chimeric	SamplingMethod	DateCode	BackTrajectory	Pressure	Temperature	Rainfall	RelHumidity	SolarFlux	Season	Month	Date	SampleDay	PM2.5	PM10	Ozone	Pollen
SC47	13847	9555	9555	9555	Spincon	SC040717	NW	984.6150685	280.9753425	0.316438356	71.25342466	183.7013699	spring	4	4/7/17	0	20	5	44	3429
HV49	38447	29133	29133	29029	Dry	HV040917	S	1004.353425	287.4205479	0.017808219	43.1369863	323.2972603	spring	4	4/9/17	2	60	13	84	722
SC414	11313	9259	9259	9259	Spincon	SC041417	SE	1007.915278	292.6263889	0.020833333	59.49027778	279.9986111	spring	4	4/17/17	10	68	0	61	3217
SC421	13662	10433	10433	10433	Spincon	SC042117	SW	1004.623288	296.2	0	63.94931507	315.7315068	spring	4	4/21/17	14	56	19	48	390
SC428	12928	10406	10406	10380	Spincon	SC042817	S	1001.383562	297.1726027	0.064383562	80.58630137	282.7945205	spring	4	4/28/17	21	56	0	31	547
SC513	48836	34620	34620	34521	Spincon	SC051317	SW	970.939726	293.5972603	0.102739726	76.67123288	276.9027397	spring	5	5/13/17	36	38	0	47	108
R518	28636	20254	20254	20219	Rain	R051817	SE	1008.320548	298.490411	0.031506849	73.45616438	342.1561644	spring	5	5/18/17	41	53	0	42	4
HV610	29344	22601	22601	22601	Dry	HV061017	SE	1005.701389	298.125	0.027777778	70.74444444	313.0708333	summer	6	6/10/17	64	55	0	58	32
R0620	26702	19752	19752	19752	Rain	R062017	SW	996.1520548	299.0863014	0.105479452	73.36712329	290.3849315	summer	6	6/20/17	74	31	12	30	0
HV0627	21607	15345	15345	15345	Dry	HV062717	E	905.9630137	289.9835616	0.001369863	57.97671233	333.5561644	summer	6	6/27/17	81	37	14	61	20
R0717	18432	14053	14053	14053	Rain	R071717	SW	983.2657534	298.0753425	0.126027397	75.13150685	272.5808219	summer	7	7/17/17	101	35	11	67	0
HV0720	11896	9113	9113	9113	Dry	HV072017	NE	948.1054795	299.4068493	0	58.76849315	340.6452055	summer	7	7/20/17	104	64	20	80	0
R721	53031	38708	38708	38517	Rain	R072117	NE	954.0972603	300.1712329	0.020547945	59.79863014	338.6547945	summer	7	7/21/17	105	66	22	122	8
R722	46627	35338	35338	34775	Rain	R072217	N	978.7575342	300.0164384	0.132876712	68.88630137	329.1972603	summer	7	7/22/17	106	62	19	71	9
HV0819	24815	19267	19267	19267	Dry	HV081917	N	954.469863	297.1958904	0.001369863	63.55068493	304.0520548	summer	8	8/19/17	134	58	19	119	0
R0831	11868	9731	9731	9731	Rain	R083117	NE	986.4068493	294.5164384	0.235616438	82.53835616	184.5739726	summer	8	8/31/17	146	32	13	24	0
R910	62276	44279	44279	44246	Rain	R091017	NE	981.8410959	287.8013699	0.032876712	67.86438356	258.9986301	fall	9	9/10/17	156	28	12	38	6
HV0913	32300	25027	25027	25012	Dry	HV091317	SW	985.690411	292.6260274	0	69.64657534	221.1931507	fall	9	9/13/17	159	49	18	43	0
HV0915	26500	20611	20611	20611	Dry	HV091517	NE	932.7123288	292.139726	0.006849315	75.22465753	195.2383562	fall	9	9/15/17	161	70	26	87	0
HV0925	39257	31190	31190	31118	Dry	HV092517	NE	913.4767123	293.4534247	0	63.37123288	220.5287671	fall	9	9/25/17	171	50	18	71	0
HV1005	25393	18804	18804	18804	Dry	HV100517	NE	986.5876712	292.1191781	0	60.55616438	238.6178082	fall	10	10/5/17	181	45	17	47	0
R1008	30951	24919	24919	24881	Rain	R100817	SE	1005.361111	299.0402778	0.45	84.46527778	193.5527778	fall	10	10/8/17	184	30	14	19	0
HV1025	21868	16846	16846	16846	Dry	HV102517	NW	968.7547945	283.430137	0	44.88630137	198.8671233	fall	10	10/25/17	201	25	11	34	0
R1028	24618	19676	19676	19676	Rain	R102817	S	1002.00274	289.2643836	0.161643836	54.15616438	166.269863	fall	10	10/28/17	204	31	11	31	0
HV1030	16286	12772	12772	12772	Dry	HV103017	SW	988.4465753	281.4410959	0	58.65479452	195.769863	fall	10	10/30/17	206	37	12	37	0
HV1102	12539	10343	10343	10343	Dry	HV110217	S	999.5638889	290.7333333	0.022222222	65.25972222	182.5722222	fall	11	11/2/17	209	60	18	38	0
R1108	29712	24346	24346	24346	Rain	R110817	NW	986.6438356	290.5643836	0.205479452	86.41917808	105.809589	fall	11	11/8/17	215	33	13	8	0
R1118	17833	14394	14394	14394	Rain	R111817	SE	1000.816438	287.5410959	0	63.27123288	170.9150685	fall	11	11/18/17	225	75	0	34	1
HV1120	19954	13873	13873	13873	Dry	HV112017	NE	968.0438356	279.2520548	0.071232877	62.57671233	132.9410959	fall	11	11/20/17	227	40	13	27	0
HV1121	31963	25394	25394	25277	Dry	HV112117	N	975.0191781	279.7753425	0.01369863	65.25342466	130.1219178	fall	11	11/21/17	228	48	18	27	0
R1215	33699	25825	25825	25825	Rain	R121517	NW	985.6027397	276.5109589	0	60.75753425	123.5219178	winter	12	12/15/17	252	21	11	20	0
S0116	20429	14086	14086	14073	Snow	S011617	S	1009.278082	287.5150685	0.012328767	63.13972603	139.6753425	winter	1	1/16/18	284	59	0	26	0
R0122	12741	9315	9315	9315	Rain	R012218	S	1003.708219	283.0630137	0	61.13013699	151.4849315	winter	1	1/22/18	290	56	0	23	3
HV0205	15913	12428	12428	12428	Dry	HV020518	NW	990.5109589	280.2452055	0.453424658	70.50821918	139.1068493	winter	2	2/5/18	304	49	16	21	1
R0207	14001	11530	11530	11530	Rain	R020718	S	1003.672603	285.9753425	0.283561644	72.39315068	160.8506849	winter	2	2/7/18	306	30	11	18	42
HV0219	21446	17308	17308	17308	Dry	HV021918	SE	1009.89726	291.5643836	0	81.13561644	138.9410959	winter	2	2/19/18	318	52	16	27	20
R0225	26502	22057	22057	21885	Rain	R022518	S	1003.238356	292.5424658	0.147945205	78.77671233	176.3657534	winter	2	2/25/18	324	34	23	30	70
R0301	19030	15110	15110	15110	Rain	R030118	SW	1007.042466	291.6041096	0.021917808	79.90547945	168.4273973	spring	3	3/1/18	328	30	15	35	56
R0306	20499	16194	16194	16040	Rain	R030618	SE	1003.733333	286.8069444	0.202777778	48.43472222	195.3041667	spring	3	3/6/18	333	59	21	37	81
HV0307	17173	12797	12797	12797	Dry	HV030718	NW	972.6575342	273.5232877	0.245205479	72.81369863	128.709589	spring	3	3/7/18	334	53	18	38	111
HV0312	12860	9500	9500	9500	Dry	HV031218	W	991.0520548	297.5109589	0	47.93150685	345.2315068	spring	3	3/12/18	339	21	7	43	5
R0329	37741	28181	28181	28126	Rain	R032918	SE	1008.484932	291.2547945	0	65.63835616	271.8232877	spring	3	3/29/18	356	53	19	40	1573
HV0402	17137	13292	13292	13234	Dry	HV040218	SW	982.8486111	291.9833333	0.173611111	87.31111111	228.8875	spring	4	4/2/18	360	66	27	61	1217
HV0418	41538	31546	31546	31546	Dry	HV041818	NW	995.8671233	288.3041096	0.010958904	63.23972603	270.230137	spring	4	4/18/18	376	45	16	64	1648
R0422	23371	19050	19050	18756	Rain	R042218	E	1016.19589	288.0753425	0.04109589	59.55205479	255.7328767	spring	4	4/22/18	380	33	13	38	681
R0426	20218	14896	14896	14896	Rain	R042618	NW	981.8643836	286.0794521	0.120547945	82.98082192	179.0246575	spring	4	4/26/18	384	35	15	29	631
HV0501	26623	20204	20204	20204	Dry	HV050118	SE	1000.975342	293.0219178	0	44.71232877	347.2027397	spring	5	5/1/18	389	61	22	87	103
HV0503	13210	9357	9357	9357	Dry	HV050318	S	1006.697222	294.6513889	0	59.41111111	331.3	spring	5	5/3/18	391	56	21	61	108
HV0509	16586	12869	12869	12869	Dry	HV050918	W	992.5493151	295.4589041	0	51.13972603	280.0958904	spring	5	5/9/18	397	58	23	97	104
R0516	43333	33979	33979	33832	Rain	R051618	SE	998.2082192	294.8479452	0.587671233	85.23561644	148.7835616	spring	5	5/16/18	404	20	14	32	41
R0529	15724	11873	11873	11873	Rain	R052918	SE	993.6328767	297.6191781	0.287671233	81.15753425	255.6246575	spring	5	5/29/18	417	23	13	24	10
R0601	15202	10930	10930	10930	Rain	R060118	SW	996.6356164	299.069863	0.397260274	80.64794521	283.5356164	summer	6	6/1/18	420	34	20	37	14
HV0604	45360	36852	36852	36852	Dry	HV060418	NW	982.2945205	294.8767123	0.079452055	63.4739726	340.0712329	summer	6	6/4/18	423	34	17	101	37
HV0619	51543	42129	42129	42129	Dry	HV061918	NW	985.1123288	298.1164384	0.142465753	75.13835616	323.5945205	summer	6	6/19/18	438	58	25	77	13
R0621	49645	37705	37705	37705	Rain	R062118	W	988.0356164	299.209589	0.098630137	69.09589041	333.3013699	summer	6	6/21/18	440	63	34	31	12
R0628	16470	13181	13181	13181	Rain	R062818	SW	997.3438356	300.5260274	0.542465753	73.25890411	289.2534247	summer	6	6/28/18	447	35	19	31	32

Figure 5 shows an Emperor plot of the Bray-Curtis distance matrix color-coded by sampling method. The Emperor plot is a graphical representation of the distance matrix, used to observe "clustering" of samples based on an external variable. Sampling method, season, and back trajectory were chosen as clustering variables because they had the highest R^2 value of any environmental variable, according to the Adonis analysis. In Figure 5, distinct clusters can be seen of dry-air Hi-Vol samples (in red) and rain samples (in blue). The results showed that the communities for each of these subsets of the community were distinct e.g., had species that appear exclusively in one subset. Figure 6 shows an Emperor plot of the Bray-Curtis distance matrix color-coded by the season of the sample. In Figure 6, the clearest cluster pattern what the separation of the Fall samples (in red), which were grouped toward the 0 point on axis 2. Figure 7 shows an Emperor plot color-coded by back-trajectory. No distinct clusters are immediately evident by visual inspection in Figure 7.

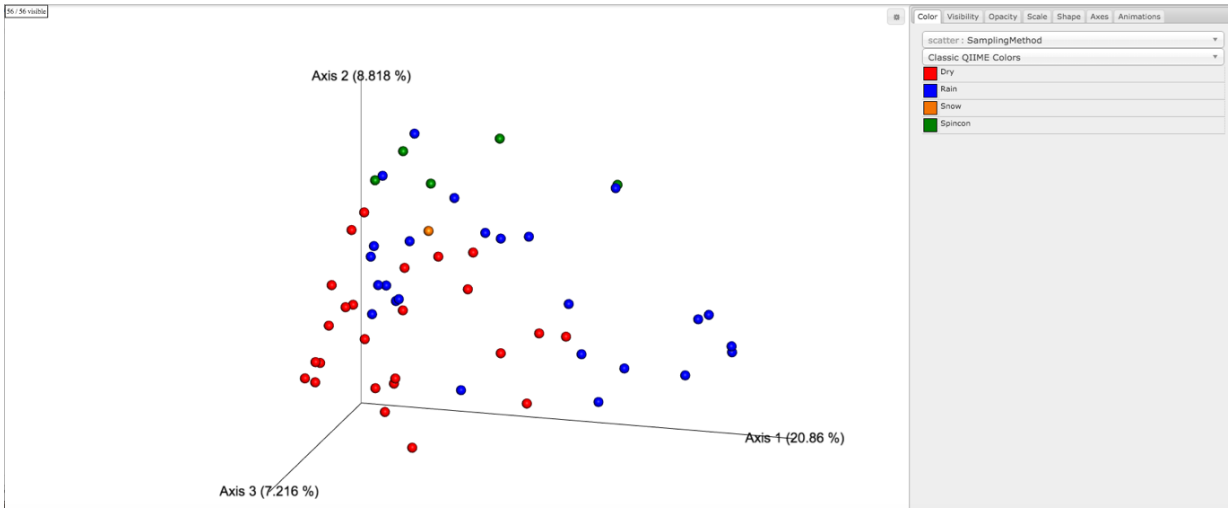


Figure 5: Emperor Plot Categorized by Sampling Method

Each point on the graph represents a different sample, and the samples are color coded by sampling method per the legend in the top right of the figure. Similarly, the following figures are color coded by the legend in the top right of each figure.

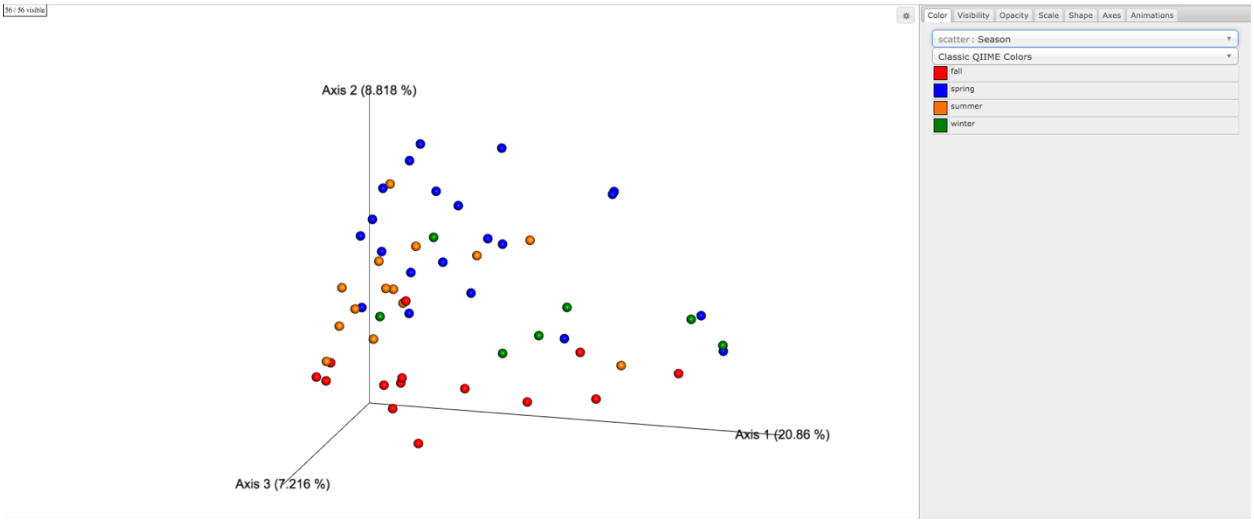


Figure 6: Emperor Plot Categorized by Season

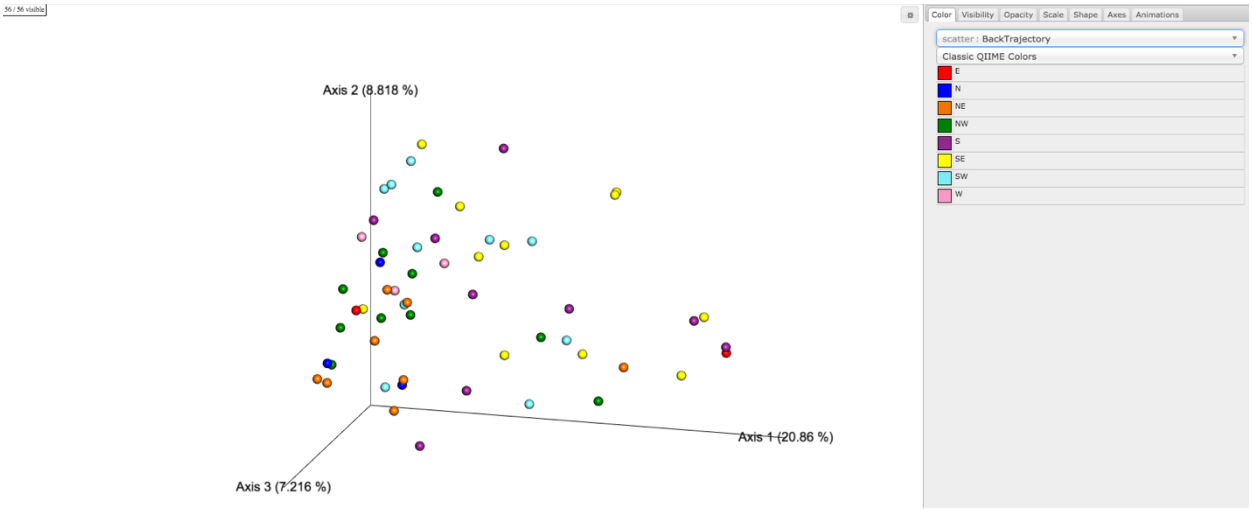


Figure 7: Emperor Plot Categorized by Back Trajectory

The airborne fungal community observed across all samples is summarized by genus in Table 3. The Qiime 2 classification results indicate that the most common ecological niche of fungi identified in the samples is as a plant pathogen. Almost all of the most genera identified interact with trees and leaves, either dead or alive. Very few of the genera contain identified species which pose a direct pathogenic or allergenic threats to humans. Approximately 26% of the representative sequences were not able to be classified at the genus level, and approximately 55% of the representative sequences were unable to be classified at the species level. The community is not dominated by certain a species or genus, each sample contains many genera. Cladosporium species make up 11.68% of each sample on average, and Stereum species make up 6.59%. There are 81 distinct genera that constitute between 0.1% and 0.5% of each sample on average, and 23 genera that constitute between 0.5% and 4% of each sample on average.

Table 3: Fungal Community Summary

Genus	Average Relative Abundance (all samples)	Present Species with Threat to Humans	Individuals at Risk	Environmental Preferences	Ecological Niche
unidentified	**25.56%**	N/A	N/A	N/A	N/A
Cladosporium	**11.68%**	C. Sphaerospermum- 0.76%	Immunocompromised	Salt-tolerant, common in atmosphere	Plant Pathogens
Stereum	**6.59%**	N/A	N/A	Deadwood, hardwood, dead leaves	Decay wood
Trametes	**3.74%**	N/A	N/A	Deadwood, hardwood, dead leaves	Decay wood
Alternaria	**3.57%**	A. Solani- 0.03%	Allergenic	Moisture, common in atmosphere	Plant Pathogens
Exidia	**1.96%**	N/A	N/A	Deadwood, hardwood, dead leaves	Decay wood
Curvularia	**1.78%**	C. Lunata- 0.48%	Immunocompromised	Wet surfaces, 24-30°C	Molds, facultative pathogens
Periconia	**1.61%**	N/A	N/A	Sac fungi	Plant Pathogens
Baeospora	**1.59%**	N/A	N/A	Coniferous forests	Mushroom producer
Aureobasidium	**1.39%**	A. Pullalans- 1.34%	Chronically Exposed	Adaptable to stressful conditions	Plant Endophyte
Ustilago	**1.03%**	N/A	N/A	Moisture	Grass Parasite
Hymenochaetopsis	**1.03%**	N/A	N/A	Trees, dead branches	Plant Pathogens
Sebipora	**1.02%**	N/A	N/A	Burned wood	White Rot
Rigidoporus	**0.81%**	N/A	N/A	Trees, wood	Plant Pathogens
Incrucipulum	**0.71%**	N/A	N/A	Leaves	Plant Pathogens
Phlebia	**0.68%**	N/A	N/A	Deadwood, conifers	Crust fungi, white rot
Mycena	**0.64%**	N/A	N/A	Deadwood, hardwood	Saprotrophic mushrooms
Peniophora	**0.63%**	N/A	N/A	Deadwood, hardwood	Plant Pathogens
Pseudomicrostroma	**0.61%**	N/A	N/A	Moisture	Plant Pathogens
Euteratosphaeria	**0.57%**	N/A	N/A	Indonesia, terrestrial	Plant Pathogens
Trichaptum	**0.56%**	N/A	N/A	Trees, wood	Mushroom producer
Ramularia	**0.56%**	N/A	N/A	Leaves	Plant Pathogens

For Blastn analysis, the reads were searched against the entire UNITE database in order to identify their closest matching fungal species representative. An example of the reads in .fasta format from sample R0122 is shown in Figure 8. Each line displays a paired read from the sample; there were 12,741 reads in this sample.

Figure 8: Sample R0122's .fasta file


R0122.fasta

```
>M01562:62:000000000-C7TNF:1:1101:19441:1926 1:N:0:GTCTGCTA+CTACTATA
TTAAGTTCAGCGGGTATTCCTACCTGATCCGAGGTCAAATTTTCAGAAGTTGGGGGTTTTACGGCAGGGCCGCGCCAGTACGATCCAAAGCGAGGGTTTTACTACTGCGCTCGGAGTCCTGGCGAGCTCG
CCACTGAATTTCAGGGCCTACCGTCCTGTAAAACAGTCAGGTAGTGCCCCAACACCAAGCTAGGCTTGAGGGTTGAAATGACGCTCGAACAGGCATGCCCTCTGGAATACCAGAGGGCGCAATGTGCGTT
>M01562:62:000000000-C7TNF:1:1101:13133:2041 1:N:0:GTCTGCTA+CTACTATT
TTAAGTTCAGCGGGTAGTCCTACCTGATTTGAGGTCAGATGTCATAAGGTTTGTCCAAAAAGGACGGTTAGAAGCTCGCCAAACGCTTCACGGTCACAGCGTAGATAATTATCACACTGAGAGCCGATCC
GCACGGAATCAAGCTAATATATTCAAGAGGAGCCGATCCCGAAAGGGACCAGCAAGCCTCCAAATCCAAGCCCATACATCACAAGGATTTATGGGTTGAGAATTCCATGACACTCAAACAGGCATGCTCC
TCGGAATACCAAGGAGCGCAAGGTGCGTT
>M01562:62:000000000-C7TNF:1:1101:16033:2047 1:N:0:GTCTGCTA+CTACTATA
TTAAGTTCAGCGGGTAGTCCCACCTGATTTGTGGTCAAAGTTCGATGAAAGCTGTCCTTTCGGACGGTTAGAAAGCGCGTCTCCACAGAAGCAAGCGCACCACAGCGTAGATAATTATCACACCGAGGCG
CACGTCGCAACAAGACGCACTAATGCATTTGAGAGGAGCCGAGCGGTGAGGCCCGGCAAAAAAACCTCCAAGTCCAAGCCACTGCTTCATTCATAAAAAATGAAGAGGGTTGAGAATTTCACGACACTCA
AACAGGTGTGCCCTTCGGAATACCAAAGGGCGCAAGGTGCGTT
>M01562:62:000000000-C7TNF:1:1101:16867:2368 1:N:0:GTCTGCTA+CTACTATA
TTAAGTTCAGCGGGTAGTCCTACCTGATTTGAGGCCAAAATAGAAAGTTCGCCATGACAAGAATGACATGGACGAGATTCGAGAGCAGGCTTCTTCTTACGAGAGTTTTCAATGTTACTTAACAACTCTG
CTGAAGCCCAGCTAATTCATTTAAGGCAAGCTTGTGTTTGTAGACCAAACCGCAAGCAGGCCCAAGGACCTGACTCGTTCTATCTTTGTTAAAAAATAAAAGAGCAGAGATATTCAAGACTCTCAAACAG
GCATGCTCCAAGGAATACCAAGGAGCGCAAGGTGCGTT
>M01562:62:000000000-C7TNF:1:1101:12025:2456 1:N:0:GTCTGCTA+CTACTATA
TTAAGTTCAGCGGGTAATCCTACCTGATTTGAGGTCAAAGTTTAAAGTTGTATAATTGGAAGCTGACATTTGCTTCACAGCTATAGTAGATATCTATCAAACTAAAGCTTATCTGCAAAGGATACAAGCT
AATACATTTAAGAGAAGCCAACTCAAAGCCAGCAACTCCAAAGTCCAAAGCCTTTTTAAAGGATTGATGATTTAATGACACTCAAACAGGCATGCTCCTCGGAATACCAAGGAGCGCAAGGTGCGTT
>M01562:62:000000000-C7TNF:1:1101:17349:2527 1:N:0:GTCTGCTA+CTACTATA
TTAAGTTCAGCGGGTAGTCCTACCTGATTTGAGGTCAGATTGTCAATGCCTATGTTACTTATAGTAAGTAACACAGTTTTTGTCTCGCAAAGGAGACGGTTAGAAGCACGGACACAATCATACTTCACGA
CCACGGCGTAGACGTTTATCACACCGAAGACGATCCGTAAAGGTTCGCACTGATACGTTTCAGAGGAGTCGATATTCTTTGCAGAGCCGACACGACCTCCAAGTCCAAGCCTTTGAGAACCAATAAAAGC
TTTCAAAGGTTGAGAATTCCATGACACTCAAACAGGCATGCTCCTCGGAATACCAAGGAGCGCAAGGTGCGTT
>M01562:62:000000000-C7TNF:1:1101:17627:2562 1:N:0:GTCTGCTA+CTACTATA
TTAAGTTCAGCGGGTATTCCTACCTGATCCGAGGTCAAACCTATAAAAATTAGGGGTTTAGTGGCCAGATGCCAGCGTTAACACTACTAAAGCGAGAATAATTACTACGCTTAGGGTCTCCGCGACTCCG
CCAATATATTTAAGGAGTTATCCCGCAACTGCAGGGTAAGCTCCCAACGCCAAGCAAATAGGGCTTGATGGTCGAAATGACGCTCGAACAGGCATGCCCACTAGAATACTAATGGGCGCAATGTGCGTT
>M01562:62:000000000-C7TNF:1:1101:14705:2567 1:N:0:GTCTGCTA+CTACTATA
TTAAGTTCAGCGGGTATCCCTACCTGATCCGAGGTCAAAGTAAAAAAGGGGGCTTGATGGGAGGAGACGGGCTGGAGACGTAGCGCAAATGTGCTGCGCACGAAGCCAACTCGCCGGCTGCCAATGAATT
TGAGGCGAGTCCACGCGCAGAGGCGGAACAAACACCCAACACCAAGCAGAGCTTGAAGGTTTAAATGACGCTCGAACAGGCATGCCCAACGGAATACCGAAGGGCGCAATGTGCGTT
>M01562:62:000000000-C7TNF:1:1101:16457:2605 1:N:0:GTCTGCTA+CTACTATA
TTAAGTTCAGCGGGTAGTCCTACCTGATTTGAGGTCAGATGTTAAAAAGTTGTCCTAATGGACGGTTAGAAGCTCGCCAAAACACTTCACGGTCACAGCGTAGACAATTATCACACTGAGAGCCGATCCG
TACGGAATCGAGCTAATGCATTCAAGAGGAGCCGACCAACAAGGGCCAGCAAGCCTCCAAGTCCAAGCTTATAGATCACAAGGATTTATAAGTTGAGAATTCCATGACACTCAAACAGGCATGCTCCTCG
GAATACCAAGGAGCGCAAGGTGCGTT
>M01562:62:000000000-C7TNF:1:1101:15621:2610 1:N:0:GTCTGCTA+CTACTATA
TTAAGTTCAGCGGGTAGTCCTACCTGATTTGAGGTCAGATGTTAAAAAGTTGTCCTAATGGACGGTTATAAGCTCGCCAAAACACTTCACGGTCACAGCGTAGACAATTATCACACTGAGAGCCGATCCG
```

Each read in each sample was compared (Blastn) to every representative sequence in the UNITE universal ITS2 fungal database using a 97% nucleotide identity threshold for a match. Table 4 displays the first 12 representative sequences of the Blastn table for sample R0122 as an example; each line represents the best match for each input read (the 12,741 reads from the .fasta file in Figure 8). An important column in this table is the fourth column: the alignment length L of each match. This indicates the length of overlap between the input read and the representative sequence in the UNITE database; the length should cover the full, or almost full, length of the query read for the match to be a reliable match. For this experiment, the average read length was 263 basepairs, so only matches with greater than 200 basepairs of overlap were considered to be reliable.

Table 4: Subset of Blastn Output Table for Sample R0122

qseqid	sseqid	pident	length	mismatch	gapopen	qstart	qend	sstart	send	evalue	bitscore	ppos	qlen	slen
M01562:62:000000000-C7TNF:1:1101:19441:1926	Valsaceae_sp|KJ739458|SH1540638.08FU|reps|k__Fungi;p__Ascomycota;c__Sordariomycetes;o__Diaporthales;f__Valsaceae;g__unidentified;s__Valsaceae_sp	98.846	260	3	0	1	260	597	338	8.06E-131	464	98.85	260	617
M01562:62:000000000-C7TNF:1:1101:13133:2041	Lenzites_betulina|UDB015755|SH1565948.08FU|refs|k__Fungi;p__Basidiomycota;c__Agaricomycetes;o__Polyporales;f__Coriolaceae;g__Lenzites;s__Lenzites_betulina	99.308	289	2	0	1	289	605	317	1.47E-148	523	99.31	289	876
M01562:62:000000000-C7TNF:1:1101:16033:2047	Stereum_complicatum|MF161283|SH1575990.08FU|reps|k__Fungi;p__Basidiomycota;c__Agaricomycetes;o__Russulales;f__Stereaceae;g__Stereum;s__Stereum_complicatum	99.34	303	1	1	1	303	641	340	9.19E-156	547	99.34	303	645
M01562:62:000000000-C7TNF:1:1101:16867:2368	Exobasidium_rostrupii|KR262425|SH1514445.08FU|reps|k__Fungi;p__Basidiomycota;c__Exobasidiomycetes;o__Exobasidiales;f__Exobasidiaceae;g__Exobasidium;s__Exobasidium_rostrupii	94.966	298	3	5	1	298	548	263	1.56E-128	457	94.97	298	548
M01562:62:000000000-C7TNF:1:1101:12025:2456	Hygrophorus_roseobrunneus|EF417821|SH1577731.08FU|reps|k__Fungi;p__Basidiomycota;c__Agaricomycetes;o__Agaricales;f__Hygrophoraceae;g__Hygrophorus;s__Hygrophorus_roseobrunneus	98.444	257	3	1	1	257	540	285	6.20E-127	451	98.44	257	1124
M01562:62:000000000-C7TNF:1:1101:17349:2527	Phlebia_tremellosa|UDB011413|SH1576304.08FU|refs|k__Fungi;p__Basidiomycota;c__Agaricomycetes;o__Polyporales;f__Meruliaceae;g__Phlebia;s__Phlebia_tremellosa	99.099	333	3	0	1	333	686	354	2.77E-171	599	99.1	333	1000
M01562:62:000000000-C7TNF:1:1101:17627:2562	Diatrypaceae_sp|AY462564|SH1614261.08FU|reps|k__Fungi;p__Ascomycota;c__Sordariomycetes;o__Xylariales;f__Diatrypaceae;g__unidentified;s__Diatrypaceae_sp	99.228	259	2	0	1	259	556	298	6.20E-132	468	99.23	259	564
M01562:62:000000000-C7TNF:1:1101:14705:2567	Sporormiaceae_sp|JX978239|SH1645881.08FU|reps|k__Fungi;p__Ascomycota;c__Dothideomycetes;o__Pleosporales;f__Sporormiaceae;g__unidentified;s__Sporormiaceae_sp	98.785	247	3	0	1	247	498	252	1.28E-123	440	98.79	247	519
M01562:62:000000000-C7TNF:1:1101:16457:2605	Trametes_versicolor|UDB011614|SH1565939.08FU|refs|k__Fungi;p__Basidiomycota;c__Agaricomycetes;o__Polyporales;f__Coriolaceae;g__Trametes;s__Trametes_versicolor	99.65	286	1	0	1	286	595	310	1.45E-148	523	99.65	286	899
M01562:62:000000000-C7TNF:1:1101:15621:2610	Trametes_versicolor|UDB011614|SH1565939.08FU|refs|k__Fungi;p__Basidiomycota;c__Agaricomycetes;o__Polyporales;f__Coriolaceae;g__Trametes;s__Trametes_versicolor	98.601	286	3	1	1	285	595	310	5.25E-143	505	98.6	285	899
M01562:62:000000000-C7TNF:1:1101:16100:2648	Trametes_versicolor|UDB011614|SH1565939.08FU|refs|k__Fungi;p__Basidiomycota;c__Agaricomycetes;o__Polyporales;f__Coriolaceae;g__Trametes;s__Trametes_versicolor	99.65	286	1	0	1	286	595	310	1.45E-148	523	99.65	286	899
M01562:62:000000000-C7TNF:1:1101:20218:2791	Ascomycota_sp|JF449722|SH1552165.08FU|reps|k__Fungi;p__Ascomycota;c__unidentified;o__unidentified;f__unidentified;g__unidentified;s__Ascomycota_sp	96.386	249	9	0	1	249	503	255	1.02E-114	411	96.39	249	1100

These tables were then analyzed further in order to find matches to putative fungal pathogens. A list of 103 putative fungal pathogens and allergens was curated as a reference database, and a text sorting function was used to find every instance of these pathogens in the Blastn tables. The text sorting function also allowed alignment length L and percent identity to be used as a filter for match quality, with only matches with L>200 basepairs (bp) and >97% nucleotide sequence identity being considered matches with high confidence for species identification. The list of 103 putative pathogens is given in Table 5.

Table 5: List of Putative Fungal Pathogens and Allergens

No.	Genus	Species	No.	Genus	Species	No.	Genus	Species
1	alternaria	metachromatica	36	alternaria	tenuissima	71	fusarium	proliferatum
2	alternaria	argyranthemi	37	apiospora	montagnei	72	Histoplasma	capsulatum
3	alternaria	carotiincultae	38	aspergillus	fumigatus	73	malassezia	furfur
4	alternaria	brassicola	39	aspergillus	flavus	74	malassezia	sympodialis
5	alternaria	alternata	40	aspergillus	niger	75	nimbya	caricis
6	alternaria	brassicae	41	aspergillus	nidulans	76	paracocciidioides	brasiliensis
7	alternaria	blumeae	42	aspergillus	oryzae	77	penicillium	citrinum
8	alternaria	capsici	43	aspergillus	versicolor	78	penicillium	oxalicum
9	alternaria	cetera	44	aspergillus	terreus	79	penicillium	brevicompactum
10	alternaria	cheiranthi	45	aureobasidium	pullulans	80	penicillium	chrysogenum
11	alternaria	cinerariae	46	beauveria	bassiana	81	pleospora	herbarum
12	alternaria	conjuncta	47	blastomyces	dermatitidis	82	pneumocystis	jirovecii
13	alternaria	crassa	48	candida	albicans	83	psilocybe	cubensis
14	alternaria	cucumerina	49	candida	boidinii	84	rhizopus	microsporus
15	alternaria	dauci	50	candida	glabrata	85	rhizopus	oryzae
16	alternaria	dumosa	51	candida	parapsilosis	86	rhodotorula	mucilaginosa
17	alternaria	eryngii	52	candida	tropicalis	87	saccharomyces	cerevisiae
18	alternaria	ethzedia	53	candida	krusei	88	sporothrix	schenckii
19	alternaria	euphorbiicola	54	candida	auris	89	stachybotrys	chartarum
20	alternaria	japonica	55	cladosporium	cladosporoides	90	stemphylium	vesicarium
21	alternaria	limoniasperae	56	cladosporium	herbarum	91	stemphylium	botryosum
22	alternaria	longipes	57	cladosporium	fulvum	92	stemphylium	callistephi
23	alternaria	macrospora	58	coccidiodes	immitis	93	talaromyces	marneffei
24	alternaria	mimicula	59	coprinus	comatus	94	thermomyces	lanuginosus
25	alternaria	mouchaccae	60	cryptococcus	neoformans	95	trichophyton	schoenleinii
26	alternaria	oregonensis	61	cryptococcus	gattii	96	trichophyton	tonsurans
27	alternaria	petroselini	62	curvularia	lunata	97	trichophyton	mentagrophytes
28	alternaria	photistica	63	embellisia	allii	98	trichophyton	rubrum
29	alternaria	porri	64	embellisia	indefessa	99	ulocladium	alternariae
30	alternaria	pseudorostrata	65	embellisia	navae-zelandiae	100	ulocladium	atrum
31	alternaria	radicina	66	embellisia	telluster	101	ulocladium	botrytis
32	alternaria	solani	67	emmonsia	pasteuriana	102	ulocladium	chartarum
33	alternaria	smyrnii	68	epicoccum	purpurascens	103	ulocladium	cucurbitae
34	alternaria	sonchi	69	fusarium	solani			
35	alternaria	tagetica	70	fusarium	culmorum			

The output of the text sorting function is a percentage of the reads in each sample that share 97% identity over a 200 bp region with one of the 103 putative pathogens. 200 was chosen as a threshold because the average input sequence length across all samples was 263 bp, with a standard deviation of 34 bp. The sequence lengths are normally distributed given the sample size of n=1,451,435, so only 3% of sequences were excluded- these were considered low quality matches.

23 of the total 103 reference pathogens or allergens were present in the sample. Table 6 describes the species that were found to match reads from the samples and a description of

disease/threat to humans, which individuals are most at risk, as well as environmental variables associated with the presence of these pathogens.

Table 6: Descriptions of Identified Blastn Pathogen Sequences (% Identity >97%, L>200 bp)

Species	Threat to Humans	Humans at Risk	Environmental Preferences	Ecological Niche
Alternaria_alternata	Upper Respiratory Infections and Asthma	Immunocompromised	Warm, humid, rainfall	Infects tomato plants
Curvularia_lunata	Phaeohyphomycosis (direct epidermal infection),Eye Infections, Allergen	Immunocompromised	Wet surfaces, 24-30°C	Mold
Stemphylium_vesicarium	Allergen- Ste v 1	Allergic individuals	Weakened plant hosts	Secondarily infects onions,
Thermomyces_lanuginosus	Endocarditis (heart inflammation)	Surgery patients	Thermophile	Compost heaps, breaks down
Alternaria_oregonensis	Allergen- Alt o 1	Allergic individuals	Pacific Northwest	Infects potato plants
Rhodotorula_mucilaginosa	Yeast Infection (meningeal, skin, ocular)	All, increased risk for central veneous catheter	Ubiquitous, plastic surfaces	Saprophytic Yeast
Aspergillus_flavus	Aspergillosis (upper respiratory infection), Liver Failure	Immunocompromised	Resilient to temperature change	Infects cereal grains, legumes, and tree nuts
Aspergillus_fumigatus	Aspergillosis (upper respiratory infection), Liver Failure	Immunocompromised	Resilient to temperature change	Compost heaps, direct pathogen
Candida_tropicalis	Candidiasis- Yeast Infection (skin, gastrointestinal, genitourinary)	Neutropenic (low white blood cells), Hospital	Tropical conditions	Fruits, Gastrointestinal tract of mammals
Candida_parapsilosis	Candidiasis- Yeast Infection (skin, gastrointestinal, genitourinary)	Neonates, Hospital patients	Human hands, devices such as prosthetics and catheters	Human commensal (hands)
Penicillium_oxalicum	Produces mycotoxins (secalonic acid D)	Consumers of tropical foods	Tropical conditions	Pearl millet rhizosphere, Infects tomato plants
Stachybotrys_chartarum	Rare cases of bleeding in the lungs (requires unnatural exposure)	Humans exposed to extremely high	Cellulose rich, moist building materials	Black mold
Rhizopus_microsporus	Nosocomial infection	Pre-term infants	Neutral pH soil, 28°C	Infects maize, sunflowers, and
Aspergillus_niger	Aspergillosis (upper respiratory infection), Fungal Ear Infection	Horticultural workers	Soil and indoor environments	Infects onions, peanuts, and grapes
Alternaria_pseudorostrata	Allergen- Alt ps 1	Allergic individuals	Middle East	Infects apricots, peaches,
Alternaria_solani	Allergen- Alt s 1	Allergic individuals	Requires host plants, abundant moisture	Infects tomato and potato
Alternaria_porri	Allergen- Alt po 1	Allergic individuals	Requires host plants, abundant moisture	Infects onions
Fusarium_solani	Cornea infections, fungal keratitis	Immunocompromised,	Soil, plant roots	Infects peas, beans, potatos
Alternaria_brassicae	Allergen- Alt br 1	Allergic individuals	Mild, wet, high rainfall	Infects broccoli and cabbage
Beauveria_bassiana	Excacerbates breathing difficulties	Immunocompromised	Used as pesticide, grows naturally in soils	Infects insects with white
Candida_glabrata	Candidiasis- Yeast Infection (skin, gastrointestinal, genitourinary)	Hospital Patients, Women	Human skin	Human commensal (mucosal tissue)
Malassezia_sympodialis	Opportunistic skin infection- atopic eczema	Individuals with atopic dermatitis conditions	Human skin	Human commesal (skin)
Saccharomyces_cerevisiae	Opportunistic lung and liver infections	CVC users,	30-35°C	Bread yeast

Figure 9 shows the relative abundance of these 23 species in each of the samples.

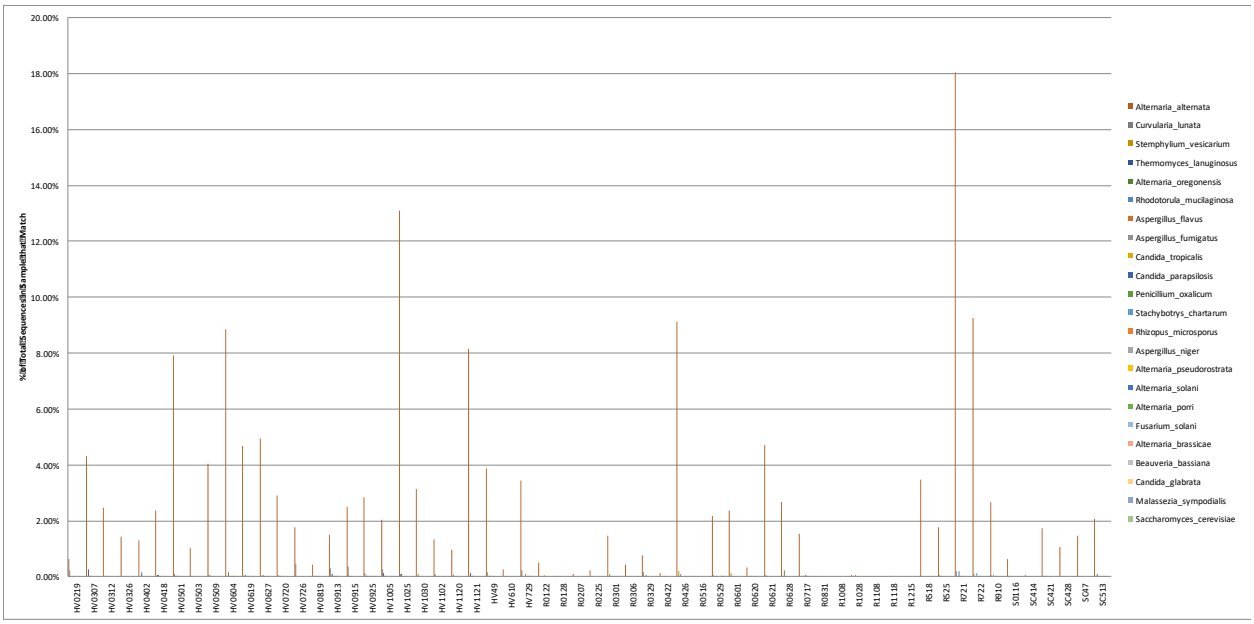


Figure 9: Relative Abundance among Reads Recovered (% Identity >97%, L>200 bp)

It is immediately evident that *Alternaria alternata* was a significant outlier in the pathogenic fraction of the community, with orders of magnitude higher relative abundance than the other 22 species found matches in the samples. Figure 10 shows the same graph with *A. alternata* removed. From Figure 10, it is clear that the pathogenic abundance (or fraction of the total reads) was at least an order of magnitude smaller without *A. alternata*, and the relative abundance of species was more consistent. That is, there are very few instances of a pathogen making up more than .2% of the reads in a sample. The vast majority of pathogenic signals were typically below the .05% mark.

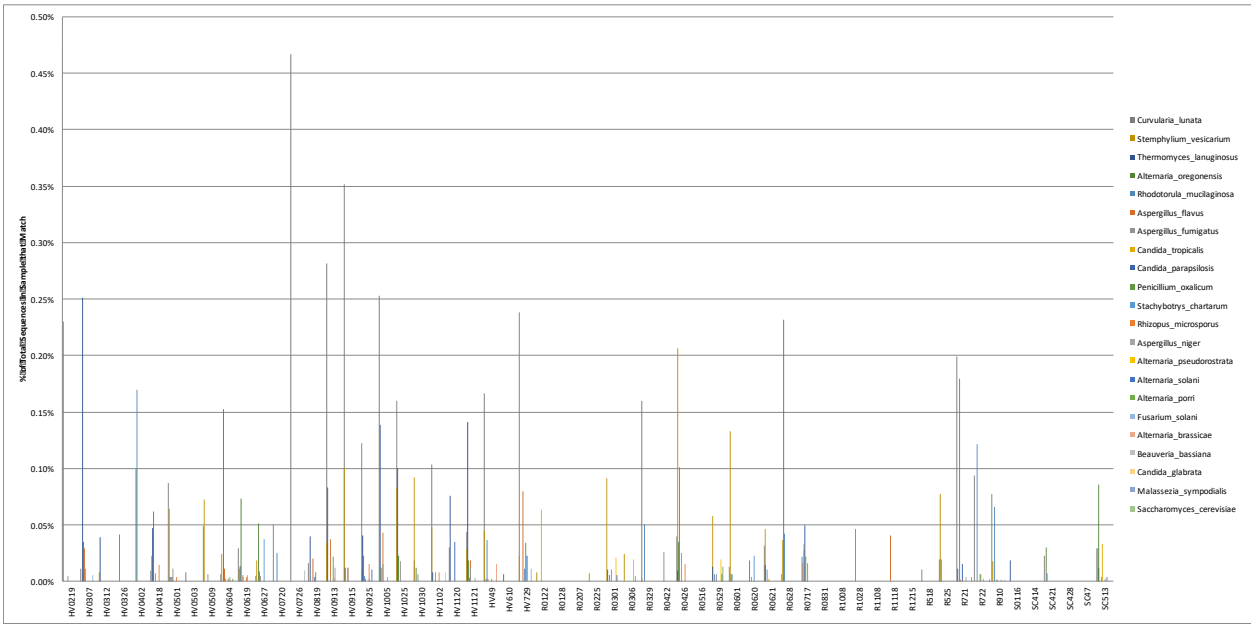


Figure 10: Relative Abundance without *A. alternata* among Reads Recovered (% Identity >97%, L>200 bp)

Using RStudio, the relative abundance data for all 23 representative sequences (including *A. alternata*) was used to create a second phylogenetic distance matrix, once again using the Bray-Curtis formula. A second Adonis analysis was conducted and is shown in Table 7. This Adonis analysis used the same 12 environmental variables as the original test, but now focusing only on the pathogenic fraction of the community. The purpose of this second Adonis analysis was to identify any external variables that had a significant influence on the presence of pathogenic fungal bioaerosols than on the community at large. The broader qualitative variables, i.e., season, sampling method, and back trajectory explained 7%, 6%, and 14%, respectively, of the variation in the pathogenic community between samples. The quantitative variables explain less of the variation, with the greatest influence represented by the $PM_{2.5}$ value (3.5%). Individually, the 12 variables explain 40.7% of the variation in the fungal communities sampled, but when taking into account linear combinations of the 12 variables, 100% of the community variation can be explained.

Table 7: Adonis Analysis for Pathogenic and Allergenic Species

	Df	SumsOfSqs	MeanSqs	F.Model	R2	Pr(>F)
Season	3	0.8023	0	0	0.07029	1
SamplingMethod	3	0.6814	0	0	0.05969	1
BackTrajectory	7	1.5836	0	0	0.13873	1
Pressure	1	0.1225	0	0	0.01073	1
Temperature	1	0.3359	0	0	0.02943	1
Rainfall	1	0.1179	0	0	0.01033	1
RelHumidity	1	0.0910	0	0	0.00798	1
SolarFlux	1	0.2914	0	0	0.02553	1
PM2Point5	1	0.4010	0	0	0.03513	1
PM10	1	0.0772	0	0	0.00677	1
Ozone	1	0.0647	0	0	0.00567	1
Pollen	1	0.0805	0	0	0.00705	1
Season:SamplingMethod	3	0.9150	0	0	0.08016	1
Season:BackTrajectory	9	2.2121	0	0	0.19380	1
SamplingMethod:BackTrajectory	7	1.6234	0	0	0.14222	1
Season:Pressure	3	0.4360	0	0	0.03819	1
SamplingMethod:Pressure	2	0.7679	0	0	0.06728	1
BackTrajectory:Pressure	4	0.4813	0	0	0.04217	1
Season:Temperature	1	0.3292	0	0	0.02884	1
Residuals	0	0.0000	-Inf		0.00000	
Total	51	11.4145			1.00000	

Figure 11 shows a brief summary of the R^2 values of each of the 12 environmental variables for both the whole community and the pathogenic fungal species fraction. Once again, season, sampling method, and back trajectory appeared to have much higher influence with changes in community composition. The trend seemed similar between both communities, which indicated the lack of major physical differences between pathogenic and non-pathogenic fungal spores in aerosolization and persistence in the atmosphere. Both pathogenic and non-pathogenic fungi reproduce in similar ways, and grow more favorably under similar conditions. There might be possible exceptions to this behavior, but pathogenic and non-pathogenic fungi differ mostly in their ecological niche and not in their reaction to ubiquitous environmental stressors or aerosolization efficiencies. Therefore, it is not

surprising that environmental factors that influence both communities are shown to influence them at a similar magnitude.

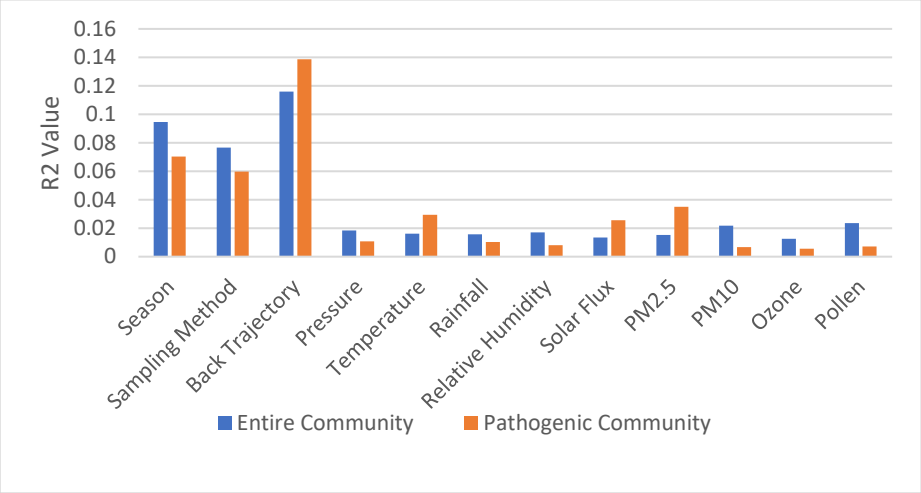


Figure 11: Summary of Individual R^2 Values from Adonis Analysis

Next, investigations into the outlier species, *A. alternata* were performed. First, an inspection of the original Qiime2 data did not confirm a presence of *A. alternata* that is equivalent to the blastn readings. Instead, Qiime2 results indicated the high abundance of *Alternaria eichhorniae*, a closely related species to *A. alternata*. Figure 12 shows the presence of *A. eichhorniae* isolated from the Qiime 2 taxonomic bar plot displayed in Figure 3. The samples with higher abundances matched approximately between each analysis, the only difference was the identified species.

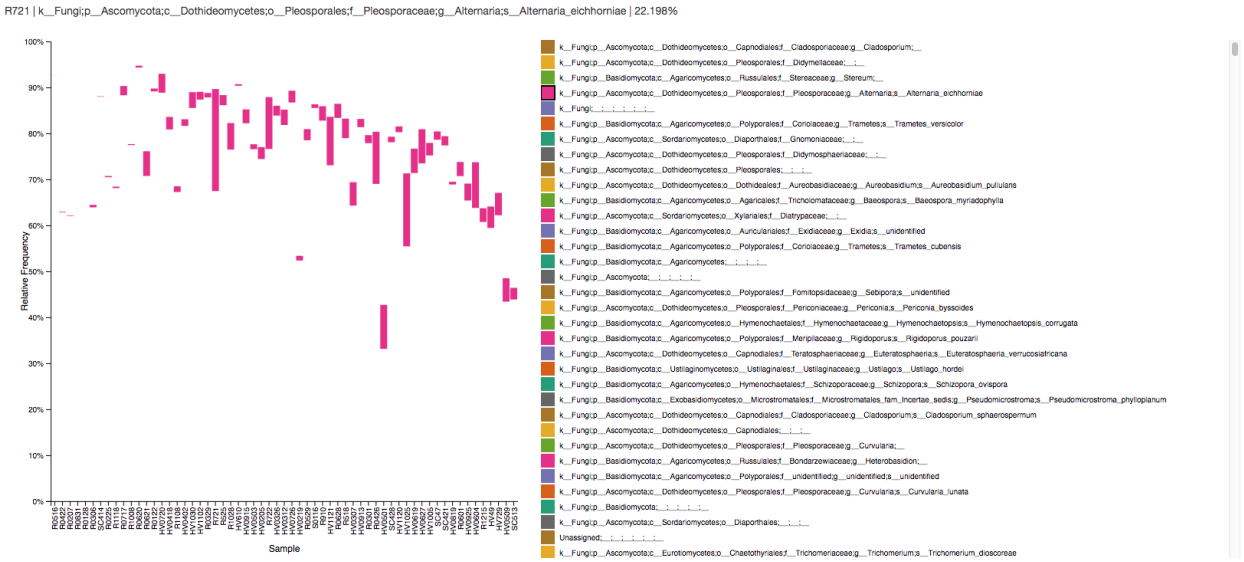


Figure 12: *A. eichhorniae* Relative Abundance in Qiime 2 Analysis

A closer look into the Blastn results of sample R721, which showed a relative high abundance of 18% *A. Alternata* in the blastn results and 22% *A. eichhorniae* in the Qiime 2 results, showed that the percent identity for matches to these two species was remarkably

close. Table 8 shows the results of the Blastn analysis, with relevant values highlighted. The highlighted rows show two examples of input reads that matched with the same number of mismatches between the two species. These results indicate that a species distinction cannot be drawn between the two representative sequences, i.e., the ITS sequence is highly conserved between the two species. Essentially, it is highly likely there is a species of the genus *Alternaria* that is present in large amounts in these samples, but the data does not convincingly determine if the species is *A. alternata* or another close relative. Because the genus *Alternaria* is known to contain opportunistic pathogens, such as *A. alternate*, it is not unreasonable to assume that a portion of the *Alternaria* species identified in the samples could indeed represent risks to individuals with pre-existing respiratory conditions.

Table 8: Blastn Results Indicating no Species Distinction Present Between *A. alternata* and *A. eichhorniae*

1	qseqid	sseqid	pident	length	mismatch	gapopen	qstart	qend	sstart	send	evalue	bitscore	ppos	qlen	slen
12	M01562:136:000000000-ANEBD:1:1101:11978:2077	Alternaria_alternata|KF465761|SH2179670	100	253	0	0	1	253	558	306	6.05E-132	468	100	253	559
14	M01562:136:000000000-ANEBD:1:1101:11978:2077	Alternaria_eichhorniae|KC146356|SH1526	99.605	253	0	1	1	253	551	300	1.01E-129	460	99.6	253	552
32	M01562:136:000000000-ANEBD:1:1101:19907:2198	Alternaria_alternata|KF465761|SH2179670	100	253	0	0	1	253	558	306	6.05E-132	468	100	253	559
34	M01562:136:000000000-ANEBD:1:1101:19907:2198	Alternaria_eichhorniae|KC146356|SH1526	99.605	253	0	1	1	253	551	300	1.01E-129	460	99.6	253	552
62	M01562:136:000000000-ANEBD:1:1101:19079:2311	Alternaria_alternata|KF465761|SH2179670	100	253	0	0	1	253	558	306	6.05E-132	468	100	253	559
64	M01562:136:000000000-ANEBD:1:1101:19079:2311	Alternaria_eichhorniae|KC146356|SH1526	99.605	253	0	1	1	253	551	300	1.01E-129	460	99.6	253	552
128	M01562:136:000000000-ANEBD:1:1101:17717:2488	Alternaria_eichhorniae|KC146356|SH1526	100	252	0	0	1	252	551	300	2.17E-131	466	100	252	552
129	M01562:136:000000000-ANEBD:1:1101:17717:2488	Alternaria_alternata|KF465761|SH2179670	99.605	253	0	1	1	252	558	306	1.01E-129	460	99.6	252	559
133	M01562:136:000000000-ANEBD:1:1101:13675:2542	Alternaria_alternata|KF465761|SH2179670	100	253	0	0	1	253	558	306	6.05E-132	468	100	253	559
135	M01562:136:000000000-ANEBD:1:1101:13675:2542	Alternaria_eichhorniae|KC146356|SH1526	99.605	253	0	1	1	253	551	300	1.01E-129	460	99.6	253	552
188	M01562:136:000000000-ANEBD:1:1101:19098:2761	Alternaria_alternata|KF465761|SH2179670	100	253	0	0	1	253	558	306	6.05E-132	468	100	253	559
190	M01562:136:000000000-ANEBD:1:1101:19098:2761	Alternaria_eichhorniae|KC146356|SH1526	99.605	253	0	1	1	253	551	300	1.01E-129	460	99.6	253	552
203	M01562:136:000000000-ANEBD:1:1101:11655:2813	Alternaria_alternata|KF465761|SH2179670	100	253	0	0	1	253	558	306	6.05E-132	468	100	253	559
205	M01562:136:000000000-ANEBD:1:1101:11655:2813	Alternaria_eichhorniae|KC146356|SH1526	99.605	253	0	1	1	253	551	300	1.01E-129	460	99.6	253	552
218	M01562:136:000000000-ANEBD:1:1101:17993:2880	Alternaria_alternata|KF465761|SH2179670	100	253	0	0	1	253	558	306	6.05E-132	468	100	253	559
220	M01562:136:000000000-ANEBD:1:1101:17993:2880	Alternaria_eichhorniae|KC146356|SH1526	99.605	253	0	1	1	253	551	300	1.01E-129	460	99.6	253	552
248	M01562:136:000000000-ANEBD:1:1101:17809:2943	Alternaria_alternata|KF465761|SH2179670	100	253	0	0	1	253	558	306	6.05E-132	468	100	253	559
250	M01562:136:000000000-ANEBD:1:1101:17809:2943	Alternaria_eichhorniae|KC146356|SH1526	99.605	253	0	1	1	253	551	300	1.01E-129	460	99.6	253	552
299	M01562:136:000000000-ANEBD:1:1101:19218:3271	Alternaria_alternata|KF465761|SH2179670	100	253	0	0	1	253	558	306	6.05E-132	468	100	253	559
301	M01562:136:000000000-ANEBD:1:1101:19218:3271	Alternaria_eichhorniae|KC146356|SH1526	99.605	253	0	1	1	253	551	300	1.01E-129	460	99.6	253	552
374	M01562:136:000000000-ANEBD:1:1101:23079:3482	Alternaria_alternata|KF465761|SH2179670	100	253	0	0	1	253	558	306	6.05E-132	468	100	253	559
376	M01562:136:000000000-ANEBD:1:1101:23079:3482	Alternaria_eichhorniae|KC146356|SH1526	99.605	253	0	1	1	253	551	300	1.01E-129	460	99.6	253	552
389	M01562:136:000000000-ANEBD:1:1101:16150:3504	Alternaria_alternata|KF465761|SH2179670	100	253	0	0	1	253	558	306	6.05E-132	468	100	253	559
391	M01562:136:000000000-ANEBD:1:1101:16150:3504	Alternaria_eichhorniae|KC146356|SH1526	99.605	253	0	1	1	253	551	300	1.01E-129	460	99.6	253	552
394	M01562:136:000000000-ANEBD:1:1101:16911:3516	Alternaria_alternata|KF465761|SH2179670	93.676	253	16	0	1	253	558	306	2.91E-105	379	93.68	253	559
396	M01562:136:000000000-ANEBD:1:1101:16911:3516	Alternaria_eichhorniae|KC146356|SH1526	93.281	253	16	1	1	253	551	300	4.88E-103	372	93.28	253	552
399	M01562:136:000000000-ANEBD:1:1101:12326:3521	Alternaria_alternata|KF465761|SH2179670	100	253	0	0	1	253	558	306	6.05E-132	468	100	253	559
401	M01562:136:000000000-ANEBD:1:1101:12326:3521	Alternaria_eichhorniae|KC146356|SH1526	99.605	253	0	1	1	253	551	300	1.01E-129	460	99.6	253	552

Discussion

Sample Flow

Of the original 60 samples for analysis, 56 were used for the Adonis analysis. The four dropped samples had too few reads to be included in the analysis. Through the Qiime 2 quality control process, a sampling depth of 9,600 reads per sample was chosen because it maximized the total number of samples with adequate coverage of the sampled diversity (near saturation of diversity, meaning very few new OTUs were recovered with more sequences after this level). This is a delicate balance because increasing the sampling depth drops samples from the analysis; a sample with 5,000 reads cannot be used if the sampling depth is 10,000. The objective of this time series is to correlate the results with metadata from different days, and a larger number of samples was preferable to a larger sampling depth. Figure 14 in the Appendix shows the alpha rarefaction curve for the Qiime 2 analysis. The asymptotic behavior indicates that increasing the sampling depth would yield only a few new OTU's.

For the Blastn pathogen analysis, some samples had to be dropped because they had no presence of any of the 103 putative pathogens. The Bray-Curtis distance matrix formula cannot

include samples with a relative abundance value of zero for all organisms. As a result, the second Adonis analysis used 52 of the original 60 samples.

Environmental Impact on Bioaerosol Communities

If a conclusion can be drawn from Figures 2 and 3, that is no snapshot of the fungal bioaerosol community looks exactly like another. The emperor plots illustrate this variability in the community as well; even when grouping by the most explanatory variables, the patterns that emerge are only slightly more organized into clusters than a random orientation.

When examining the pathogenic fraction, each environmental factor explained a similar percentage of the phylogenetic distance between samples to those for the pathogenic community, as illustrated by Figure 11. An important output of the Adonis analyses to note is the residuals. The residuals totaled zero for both iterations. This indicated at first glance that there is no variation in the community that cannot be explained by a combination of the 12 environmental variables. While this is true, it seemed that not all 12 variables were required to completely explain the community variation. Figure 15 in the Appendix explores an interesting phenomenon with R^2 called "over-explanation." When performing multivariate regression, adding more explanatory variables will always increase the Total R^2 value, as each new R^2 value must be between 0 and 1. However, the Total R^2 value cannot increase above 1, i.e. the variables cannot explain greater than 100% of the variation in the fungal communities. The Total R^2 value is equal to the sum of the R^2 values for each single variable plus the R^2 values of each linear combination of multiple variables. It is also equal to 1 minus the residual value presented by the Adonis analysis. With independent variables, the Total R^2 value will only reach 1 once every possible explanatory variable is included. With dependent variables (such as Season and Temperature), the Total R^2 value will reach 1 before every possible explanatory variable is included, because there is overlap between dependent variables' linear combination and their individual R^2 values.

The experiment shown in Figure 15 was performed to understand how many variables are necessary to first reach the 100% explanation threshold (a Total R^2 value of 1). The experiment started with 3 variables: Season, Sampling Method, and Back Trajectory. The residual value was 0.3565, meaning that the Total R^2 value is .6435. These variables and their linear combinations together explained 64.35% of the community variation. Next, Temperature was added as an explanatory variable, decreasing the residual value to 0.1314. Relative humidity was added next, and the residual value decreased to 0.01782. With only 5 of the 12 environmental variables, the total community variation explained was greater than 98%, and by adding $PM_{2.5}$ as a sixth variable, the residual value dropped to 0. This indicates that 100% of the variation in the community can be explained by only these 6 variables and their linear combinations.

Analyzing additional variables will not increase the displayed Total R^2. However, because the explanatory variables are not independent, there is overlap in the explanatory power of the linear combinations with the explanatory power of their individual components (e.g. explanatory power attributed to a linear combination of Season and Temperature will be too large due to Temperature's dependence on Season). With 6 variables at the end of the experiment

in Figure 15, only 33.3% of the total variation is explained by individual variables, without linear combinations of multiple variables. Adding more variables will increase this number (eventually to 44.1% for the entire community, and 40.7% for the pathogenic community) while keeping the Total R^2 value at 1. This is because some of the over-explanation due to dependent variables is now attributed to individual variables, as increasing the total degrees of freedom dilutes the explanatory power of any linear combination. For example, in Table 1, the explanatory power of the Season:Temperature linear combination has been reduced to 2.7% once all 12 variables have been included. This is relative to 6.2% with only 6 variables (the last output of Figure 15). Adding more variables and comparing their individual R^2 values is also useful to assess relative influence of the 12 explanatory variables, as shown in Figure 11.

Understanding the Risk of Breathing Open Air

A majority of the genera observed in the community as a whole are plant pathogens, very few of which represent allergenic risks to humans. Because Atlanta is a city surrounded by forest, with wind masses originating from all cardinal directions, the observation of many distinct wood pathogens at small relative abundance fractions is not surprising. There is an abundance of hosts for these pathogenic fungi (either live or dead trees), so there is little reason to believe certain genera will out-compete others and dominate the atmospheric community. The risk to humans from breathing ambient air is limited to a small portion of the community.

Overall, the pathogenic fraction of the community is relatively small, averaging approximately 1.4% of total reads in the sample without including reads assigned to *A. alternata* and approximately 2.8% per sample with *A. alternata* included. Using Prussin et al.'s estimation for fungal spore abundance of 10^2 to 10^3 fungal spores per m^3 in open air, an estimate can be calculated for the total fungal pathogenic exposure of the average human, who breathes in about 7.5 L per minute (10.8 m^3/day). According to the EPA, only about 7% of the average American's day is spent outdoors.

Average Value (550 spores/m^3, 7% of time outdoors, 1.4% pathogenic fraction)**:**

$$\frac{550\ fungal\ spores}{m^3\ of\ open\ air} * \frac{10.8\ m^3\ air}{day} * .07\ * 0.014\ = \frac{5.82\ pathogenic\ spores}{day}$$

Upper Bound (10^3 spores/m^3, all day outdoors, 2.8% pathogenic fraction)**:**

$$\frac{10^3\ fungal\ spores}{m^3\ of\ open\ air} * \frac{10.8\ m^3\ air}{day} * 1\ * 0.028\ = \frac{302.4\ pathogenic\ spores}{day}$$

In combination with the information in Table 5, these calculations indicate that the risk of fungal infection by the 23 pathogens found in the samples at the level of exposure described above is only significant for immunocompromised, allergenic, or hospitalized individuals. For the average American with a healthy immune system, 6 pathogenic or allergenic spores per day hardly seems a significant risk. In fact, the American Conference of Governmental Industrial Hygienists considers 100 pathogenic or allergenic spores per m^3 to be the maximum concentration considered as "low risk" for indoor air, with 100 to 1000 CFU/m^3 presenting "intermediate risk" and >1000 CFU/m^3 presenting "high risk" [20]. Assuming humans are indoors 93% of the time, and breathe 0.8 m^3 air/day, this means breathing in less than 1000

pathogenic or allergenic spores per day in a “low risk” indoor environment (greater than the outdoor upper bound).

With the 23 pathogens identified, cases of direct infection of healthy individuals were rare and involved exposure to larger concentrations of the pathogen than those present in outdoor air (i.e. cases of indoor mold exposure and nosocomial infections). At the levels of exposure calculated above, only individuals who are particularly susceptible to infection and who spend a significant amount of time outdoors may be at risk for infection from bioaerosols in the atmosphere. For example, immunocompromised individuals experiencing homelessness would be particularly at risk for fungal infection. Other factors such as exposure to damp surfaces and lack of hygienic routine could also increase risk of infection. Preventative measures can still be taken for those who spend most of their time indoors but have special susceptible conditions such as asthma. The most predominant infections from outdoor interaction with airborne fungi are of the upper respiratory system, so a simple air mask to prevent spores from entering the respiratory system could help prevent infections. For those who are allergenic, taking antihistamine medication and keeping an epinephrine injector for emergencies are viable counter-measures.

Sources of Error

Errors in bench work throughout the sampling, extracting, and amplifying processes are more likely to cause bias in results than errors after PCR amplification. Any exposure of filters, samples, and reagents to open air prior to amplification could lead to false positive signal. Sterility is still practiced rigorously before and after amplification, but there is a certain amount of exposure to indoor air that cannot be avoided. During the purification process, the concentration of magnetic beads used determines the length of amplicon that is retained. Accidental deviation from this concentration could lead to bias against amplicons of certain lengths.

A. alternata as an outlier in the fungal community brought attention to a significant source of error: a difference in identification between the Qiime 2 software and the Blastn software. Because the reads had to be re-paired outside of Qiime 2 in order to use Blastn, final results in identification differed between each method. This error has significant implications on the results, as *A. alternata* is a plant pathogen and human allergen, while *A. eichorniae* is a plant pathogen but not a human allergen.

Conclusions

When it comes to characterizing the risk of breathing open air, the data do not indicate that any specific environmental factor can significantly increase the pathogenic fraction of the fungal bioaerosol community. Instead, as is the case with the community as a whole, a combination of many factors explains the variation in (potentially) pathogenic fraction of the airborne composition. It is worth noting that the relative abundance of an individual species can vary dramatically between samples, so there may be certain days that see a localized spike in abundance of a particular threatening species. The risk of direct pathogenic infection by a fungal bioaerosol seems low in these outdoor environments, in general. Individuals with pre-existing

respiratory conditions may need to consider fungal bioaerosols as a small fraction of the ambient influences on their health. However, the risk to these individuals due to fungal bioaerosols seems to be relatively small compared to other dangerous airborne particles such as particulate matter, ozone, and pollen due to the low concentration of pathogenic and allergenic fungal bioaerosols in the atmosphere.

Future Work

Data could also be referenced against Center for Disease Control (CDC) statistics on the severity or number of asthmatic attacks in Atlanta to determine the strength of the correlation between fungal pathogen atmospheric abundance and public health. Studies could also be conducted using qPCR and viral particles collected using the slit cyclone sampler with a similar end goal in mind: to determine the risk of breathing open air. Analysis of the hydrophilicity of certain species of bacteria and fungi could establish whether airborne microbes serve as ice nuclei for the formation of clouds. Isolate-based freezing experiments are currently being run in the enve-omics lab (Lizbeth Davila, graduate student) in order to identify bacterial species that serve as effective ice nuclei. Flow cytometry could be used to determine the average size and quantity of cells in each sample. It could also be used to determine whether the membranes of cells are intact (indicating viability). This is an important distinction as some species will activate stages of dormancy in harsh conditions such as those prevailing in the atmosphere, and other collected cells will simply be dead and unable to reproduce. When studying the infectivity of pathogenic bioaerosols, membrane integrity would be a first step to determine whether they truly pose a threat to human health. Another important step would to culture isolates of the observed pathogens and observing their ability to reproduce properly. Two important environmental factors that merit future experiments based on the influence of back trajectory are elevation and location. All of these samples were taken at the same location and elevation, and if simultaneous samples were taken at slightly different locations, the impact of these variables could be analyzed. For example, the concentration of microbes at ground level could be different from the concentration on the roof. Previous literature has shown that fungal spore concentration decreases with drastic altitude increases (>2000m), so it is possible but unlikely that a change of similar magnitude would be observed between the ground and the roof [21]. As population density skyrockets in urban areas around the world, understanding the microbial communities in the atmosphere will play a critical role in limiting the spread of infection. Given the increased risk for immunocompromised individuals who spend more time outdoors, collaboration with local homeless shelters could help bolster efforts to reduce risk of infection in those who are particularly susceptible.

References

1. Verreault, D., Moineau, S. & Duchaine, C. Methods for Sampling of Airborne Viruses. *Microbiology and Molecular Biology Reviews* **72**, 413-444 (2008).
2. Groulx, N., Urch, B., Duchaine, C., Mubareka, S. & Scott, J.A. The Pollution Particulate Concentrator (PoPCon): A platform to investigate the effects of particulate air pollutants on viral infectivity. *Science of The Total Environment* **628-629**, 1101-1107 (2018).
3. Prussin Ii, A., Marr, L. & Bibby, K. *Challenges of studying viral aerosol metagenomics and communities in comparison with bacterial and fungal aerosols*, (2014).
4. Booth, T.F. et al. Detection of airborne severe acute respiratory syndrome (SARS) coronavirus and environmental contamination in SARS outbreak units. *J Infect Dis* **191**,1472-7 (2005).
5. Emery, S.L. et al. Real-Time Reverse Transcription–Polymerase Chain Reaction Assay for SARS-associated Coronavirus. *Emerging Infectious Diseases* **10**, 311-316 (2004).
6. Bing, Y., Zhang, Y.-H., Leung, N.H.L., Cowling, B.J. & Yang, Z.-F. Role of viral bioaerosols in nosocomial infections and measures for prevention and control. *Journal of Aerosol Science* **117**, 200-211 (2018).
7. Dowell, S.F. Seasonal variation in host susceptibility and cycles of certain Infectious diseases. *Emerging Infectious Diseases* **7**, 369-374 (2001).
8. DeLeon-Rodriguez, Natasha, et al. "Microbiome of the Upper Troposphere: Species Composition and Prevalence, Effects of Tropical Storms, and Atmospheric Implications." *PNAS*, National Academy of Sciences, 12 Feb. 2013, www.pnas.org/content/110/7/2575.
9. Facts About the Common Cold. (American Lung Association Scientific and Medical Editorial Review Panel, 2018).
10. Medaglia, S. SpinCon® Based Air Sampler Foot and Mouth Disease Virus (FMDV) Test Report (Sceptor Industries, Plum Island Animal Disease Center, 2001).
11. Real Time PCR: Understanding C_t. (Applied Biosystems- Life Technologies Inc., 2011).
12. WHO information for molecular diagnosis of influenza virus - update. *World Health Organization* (2017).
13. van Elden, L.J.R., Nijhuis, M., Schipper, P., Schuurman, R. & van Loon, A.M. Simultaneous Detection of Influenza Viruses A and B Using Real-Time Quantitative PCR. *Journal of Clinical Microbiology* **39**, 196-200 (2001).
14. Fabian, P., McDevitt, J.J., Lee, W.-M., Houseman, E.A. & Milton, D.K. An optimized method to detect influenza virus and human rhinovirus from exhaled breath and the airborne environment. *Journal of environmental monitoring: JEM* **11**, 314-317 (2009).
15. Ihrmark, K. et al. New primers to amplify the fungal ITS2 region--evaluation by 454-sequencing of artificial and natural communities. *FEMS Microbiol Ecol* **82**, 666-77 (2012).
16. Konopka J, Casadevall A, Taylor J, Heitman J, Cowen L. 2019. One health: fungal pathogens of humans, animals, and plants.
17. Simon-Nobbe, B., Denk, U., Poll, V., Rid, R., & Breitenbach, M. (2008). The spectrum of fungal allergy. *Int Arch Allergy Immunol, 145*(1), 58-86. doi:10.1159/00010757
18. Negron, A., DeLeon-Rodriguez, N., Waters, S. M., Ziemba, L. D., Anderson, B., Bergin, M., . . . Nenes, A. (2018). Using flow cytometry and light-induced fluorescence technique to characterizethe variability and characteristics of bioaerosols in springtime at Metro Atlanta, Georgia. Atmos. Chem. *Phys. Discuss., 2018*, 1-42. doi:10.5194/acp-2018-1073
19. Oh S., Kurt Z., Tsementzi D., Weigand M. R., Kim M., Hatt J. K., Tandukar M., Pavlostathis S. G., Spain J. C., and Konstantinidis K. T. Microbial community degradation of widely used

quaternary ammonium disinfectants. (2014). Applied and Environmental Microbiology. Oct; 80(19):5892-900.
20. Cohen, Beverly S, et al. Air Sampling Instruments for Evaluation of Atmospheric Contaminants. 8th ed., ACGIH, 1995.
21. Pace L., Boccacci L., Casilli M., Fattorini S., Temporal variations in the diversity of airborne fungal spores in a Mediterranean high altitude site (2019). Atmospheric Environment, Volume 210, Pages 166-170, ISSN 1352-2310, https://doi.org/10.1016/j.atmosenv.2019.04.059.

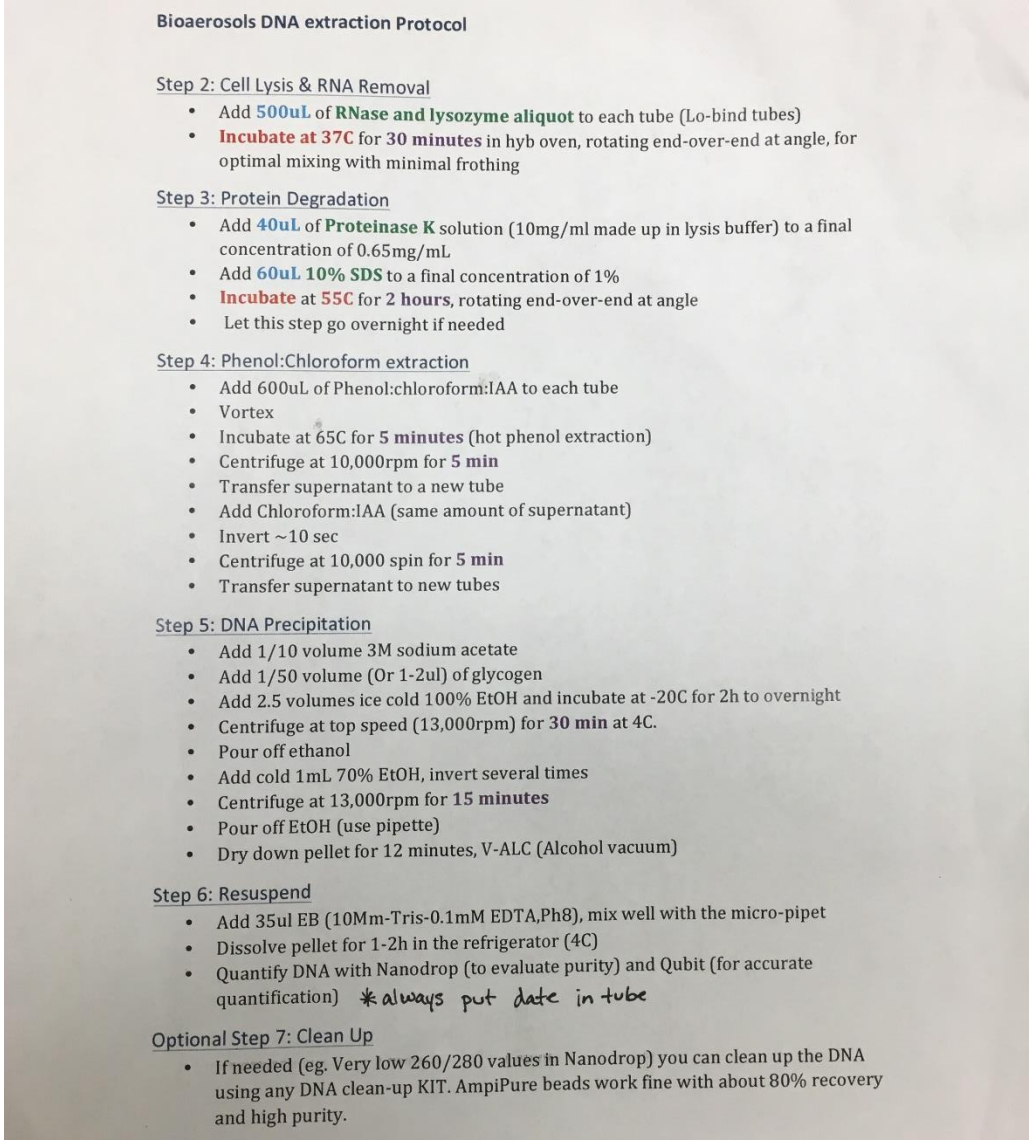
Appendix

Figure 13: High-Volume Sampler on the Roof of Ford ES&T Building

Table 9: Phosphate-Buffered Saline (PBS) Recipe

Salt	Concentration (mmol/L)	Concentration (g/L)
NaCl	137	8.0
KCl	2.7	0.2
Na_2HPO_4	10	1.44
KH_2PO_4	1.8	0.24

DNA Extraction Protocol (Developed by Lizbeth Davila in the Konstantinidis (enve-omics) Lab.



Bioaerosols DNA extraction Protocol

Step 2: Cell Lysis & RNA Removal

- Add **500uL** of **RNase and lysozyme aliquot** to each tube (Lo-bind tubes)
- **Incubate at 37C** for **30 minutes** in hyb oven, rotating end-over-end at angle, for optimal mixing with minimal frothing

Step 3: Protein Degradation

- Add **40uL** of **Proteinase K** solution (10mg/ml made up in lysis buffer) to a final concentration of 0.65mg/mL
- Add **60uL 10% SDS** to a final concentration of 1%
- **Incubate** at **55C** for **2 hours**, rotating end-over-end at angle
- Let this step go overnight if needed

Step 4: Phenol:Chloroform extraction

- Add 600uL of Phenol:chloroform:IAA to each tube
- Vortex
- Incubate at 65C for **5 minutes** (hot phenol extraction)
- Centrifuge at 10,000rpm for **5 min**
- Transfer supernatant to a new tube
- Add Chloroform:IAA (same amount of supernatant)
- Invert ~10 sec
- Centrifuge at 10,000 spin for **5 min**
- Transfer supernatant to new tubes

Step 5: DNA Precipitation

- Add 1/10 volume 3M sodium acetate
- Add 1/50 volume (Or 1-2ul) of glycogen
- Add 2.5 volumes ice cold 100% EtOH and incubate at -20C for 2h to overnight
- Centrifuge at top speed (13,000rpm) for **30 min** at 4C.
- Pour off ethanol
- Add cold 1mL 70% EtOH, invert several times
- Centrifuge at 13,000rpm for **15 minutes**
- Pour off EtOH (use pipette)
- Dry down pellet for 12 minutes, V-ALC (Alcohol vacuum)

Step 6: Resuspend

- Add 35ul EB (10Mm-Tris-0.1mM EDTA,Ph8), mix well with the micro-pipet
- Dissolve pellet for 1-2h in the refrigerator (4C)
- Quantify DNA with Nanodrop (to evaluate purity) and Qubit (for accurate quantification) *always put date in tube

Optional Step 7: Clean Up

- If needed (eg. Very low 260/280 values in Nanodrop) you can clean up the DNA using any DNA clean-up KIT. AmpiPure beads work fine with about 80% recovery and high purity.

Table 10: Fungal ITS Primer Sequences used in t.his study

Primer array	**Sequence**
ITSF.SB501	AATGATACGGCGACCACCGAGATCTACACCTACTATATATGGTAATTGGTCCTCCGCTTATTGATATGC
ITSF.SB502	AATGATACGGCGACCACCGAGATCTACACCGTTACTATATGGTAATTGGTCCTCCGCTTATTGATATGC
ITSF.SB503	AATGATACGGCGACCACCGAGATCTACACAGAGTCACTATGGTAATTGGTCCTCCGCTTATTGATATGC

ITSF.SB504	AATGATACGGCGACCACCGAGATCTACACTACGAGACTATGGTAATTGGTCCTCCGCTTATTGATATGC
ITSF.SB505	AATGATACGGCGACCACCGAGATCTACACACGTCTCGTATGGTAATTGGTCCTCCGCTTATTGATATGC
ITSF.SB506	AATGATACGGCGACCACCGAGATCTACACTCGACGAGTATGGTAATTGGTCCTCCGCTTATTGATATGC
ITSF.SB507	AATGATACGGCGACCACCGAGATCTACACGATCGTGTTATGGTAATTGGTCCTCCGCTTATTGATATGC
ITSF.SB508	AATGATACGGCGACCACCGAGATCTACACGTCAGATATATGGTAATTGGTCCTCCGCTTATTGATATGC
ITSR.SA701	CAAGCAGAAGACGGCATACGAGATAACTCTCGAGTCAGTCAGCCGTGARTCATCGAATCTTTG
ITSR.SA702	CAAGCAGAAGACGGCATACGAGATACTATGTCAGTCAGTCAGCCGTGARTCATCGAATCTTTG
ITSR.SA703	CAAGCAGAAGACGGCATACGAGATAGTAGCGTAGTCAGTCAGCCGTGARTCATCGAATCTTTG
ITSR.SA704	CAAGCAGAAGACGGCATACGAGATCAGTGAGTAGTCAGTCAGCCGTGARTCATCGAATCTTTG
ITSR.SA705	CAAGCAGAAGACGGCATACGAGATCGTACTCAAGTCAGTCAGCCGTGARTCATCGAATCTTTG
ITSR.SA706	CAAGCAGAAGACGGCATACGAGATCTACGCAGAGTCAGTCAGCCGTGARTCATCGAATCTTTG
ITSR.SA707	CAAGCAGAAGACGGCATACGAGATGGAGACTAAGTCAGTCAGCCGTGARTCATCGAATCTTTG
ITSR.SA708	CAAGCAGAAGACGGCATACGAGATGTCGCTCGAGTCAGTCAGCCGTGARTCATCGAATCTTTG
ITSR.SA709	CAAGCAGAAGACGGCATACGAGATGTCGTAGTAGTCAGTCAGCCGTGARTCATCGAATCTTTG
ITSR.SA710	CAAGCAGAAGACGGCATACGAGATTAGCAGACAGTCAGTCAGCCGTGARTCATCGAATCTTTG
ITSR.SA711	CAAGCAGAAGACGGCATACGAGATTCATAGACAGTCAGTCAGCCGTGARTCATCGAATCTTTG
ITSR.SA712	CAAGCAGAAGACGGCATACGAGATTCGCTATAAGTCAGTCAGCCGTGARTCATCGAATCTTTG

Table 11: PCR Reaction Mix Per Tube (Total Volume: 25 µL)

Component	**Volume (µL)**
Millipure Water	17.25
Accuprime 10x Reaction Mix	2.5
BSA (10 ng/µL)	1.25
Accuprime Taq DNA Polymerase	0.25
DNA Template (1:10 Dilution in EB Buffer)	1.25
Forward Primer (10 µM stock in TE Buffer)	1.25
Reverse Primer (10 µM stock in TE Buffer)	1.25

Table 12: Thermocycler Reaction Sequence

Temperature (˚C)	Time	Cycles
95	2 min	1x
95	30 sec	25x
52	30 sec	
72	1 min	
72	6 min	1x
4	hold	1x

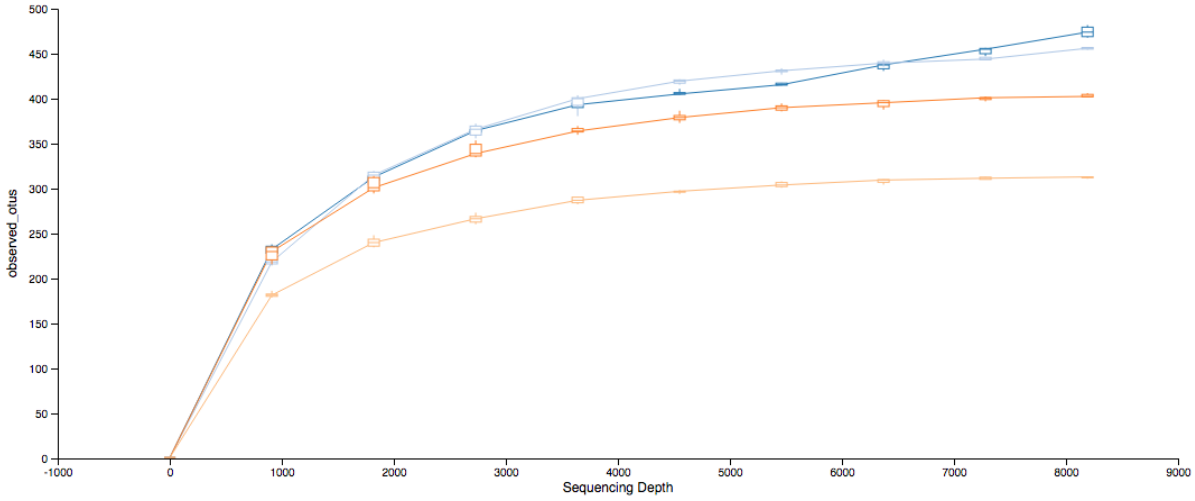


Figure 14: Alpha Rarefaction Curve Performed with Qiime 2

```
Call:
adonis(formula = bray.curtis.distance.matrix ~ Season * SamplingMethod *      BackTrajectory, data = metaforadonisFall2019wPM03)

Permutation: free
Number of permutations: 999

Terms added sequentially (first to last)

                            Df SumsOfSqs MeanSqs F.Model      R2 Pr(>F)
Season                       3    1.5926 0.53088 1.76767 0.09451  0.003 **
SamplingMethod               3    1.2912 0.43042 1.43316 0.07663  0.021 *
BackTrajectory               7    1.9536 0.27908 0.92926 0.11593  0.698
Season:SamplingMethod        3    0.6193 0.20642 0.68733 0.03675  0.975
Season:BackTrajectory       11    3.2264 0.29330 0.97662 0.19147  0.562
SamplingMethod:BackTrajectory  8    2.1610 0.27013 0.89944 0.12825  0.775
Residuals                   20    6.0066 0.30033         0.35646
Total                       55   16.8507                 1.00000
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> adonis(formula = bray.curtis.distance.matrix ~ Season * SamplingMethod * BackTrajectory * Temperature, data = metaforadonisFall2019wPM0
3)

Call:
adonis(formula = bray.curtis.distance.matrix ~ Season * SamplingMethod *      BackTrajectory * Temperature, data = metaforadonisFall2019w
PM03)

Permutation: free
Number of permutations: 999

Terms added sequentially (first to last)

                                 Df SumsOfSqs MeanSqs F.Model      R2 Pr(>F)
Season                            3    1.5926 0.53088 1.91817 0.09451  0.001 ***
SamplingMethod                    3    1.2912 0.43042 1.55518 0.07663  0.015 *
BackTrajectory                    7    1.9536 0.27908 1.00838 0.11593  0.436
Temperature                       1    0.2709 0.27086 0.97866 0.01607  0.440
Season:SamplingMethod             3    0.6428 0.21426 0.77415 0.03815  0.903
Season:BackTrajectory            11    3.2204 0.29277 1.05783 0.19112  0.325
SamplingMethod:BackTrajectory     8    2.2013 0.27516 0.99422 0.13064  0.494
Season:Temperature                3    1.0567 0.35225 1.27274 0.06271  0.099 .
SamplingMethod:Temperature        2    0.6570 0.32850 1.18695 0.03899  0.190
BackTrajectory:Temperature        5    1.5340 0.30680 1.10851 0.09103  0.258
Season:BackTrajectory:Temperature  1    0.2160 0.21597 0.78036 0.01282  0.772
Residuals                         8    2.2141 0.27676         0.13140
Total                            55   16.8507                 1.00000
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1


Call:
adonis(formula = bray.curtis.distance.matrix ~ Season * SamplingMethod *      BackTrajectory * Temperature * RelHumidity, data = metaforadoni
sFall2019wPM03)

Permutation: free
Number of permutations: 999

Terms added sequentially (first to last)

                                    Df SumsOfSqs MeanSqs F.Model      R2 Pr(>F)
Season                               3    1.5926 0.53088 1.76751 0.09451  0.066 .
SamplingMethod                       3    1.2912 0.43042 1.43303 0.07663  0.148
BackTrajectory                       7    1.9536 0.27908 0.92918 0.11593  0.643
Temperature                          1    0.2709 0.27086 0.90179 0.01607  0.583
RelHumidity                          1    0.2880 0.28804 0.95901 0.01709  0.533
Season:SamplingMethod                3    0.6461 0.21536 0.71701 0.03834  0.844
Season:BackTrajectory               11    3.2209 0.29281 0.97489 0.19115  0.588
SamplingMethod:BackTrajectory        8    2.2404 0.28005 0.93239 0.13296  0.642
Season:Temperature                   3    1.0647 0.35491 1.18163 0.06319  0.331
SamplingMethod:Temperature           2    0.6054 0.30271 1.00785 0.03593  0.526
BackTrajectory:Temperature           5    1.5058 0.30117 1.00270 0.08936  0.552
Season:RelHumidity                   3    0.9166 0.30553 1.01722 0.05439  0.534
BackTrajectory:RelHumidity           1    0.3395 0.33946 1.13021 0.02015  0.367
Temperature:RelHumidity              1    0.4031 0.40308 1.34202 0.02392  0.243
Season:Temperature:RelHumidity       2    0.2115 0.10573 0.35200 0.01255  0.995
Residuals                            1    0.3004 0.30035         0.01782
Total                               55   16.8507                 1.00000
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
Call:
adonis(formula = bray.curtis.distance.matrix ~ Season * SamplingMethod *      BackTrajectory * Temperature * RelHumidity * PM2Point5, data =
metaforadonisFall2019wPM03)

Permutation: free
Number of permutations: 999

Terms added sequentially (first to last)

                             Df SumsOfSqs MeanSqs F.Model      R2 Pr(>F)
Season                        3    1.5926       1       0 0.09451      1
SamplingMethod                3    1.2912       0       0 0.07663      1
BackTrajectory                7    1.9536       0       0 0.11593      1
Temperature                   1    0.2709       0       0 0.01607      1
RelHumidity                   1    0.2880       0       0 0.01709      1
PM2Point5                     1    0.2192       0       0 0.01301      1
Season:SamplingMethod         3    0.6270       0       0 0.03721      1
Season:BackTrajectory        11    3.2407       0       0 0.19232      1
SamplingMethod:BackTrajectory 8    2.3453       0       0 0.13918      1
Season:Temperature            3    1.0428       0       0 0.06189      1
SamplingMethod:Temperature    2    0.7088       0       0 0.04207      1
BackTrajectory:Temperature    5    1.4434       0       0 0.08566      1
Season:RelHumidity            3    0.9956       0       0 0.05908      1
BackTrajectory:RelHumidity    1    0.3017       0       0 0.01790      1
Temperature:RelHumidity       1    0.2859       0       0 0.01697      1
Season:PM2Point5              2    0.2438       0       0 0.01447      1
Residuals                     0    0.0000     Inf         0.00000
Total                        55   16.8507                 1.00000
```

Figure 15: Adonis Iteration to Indentify Number of Variables that "Overexplain" Community Variation

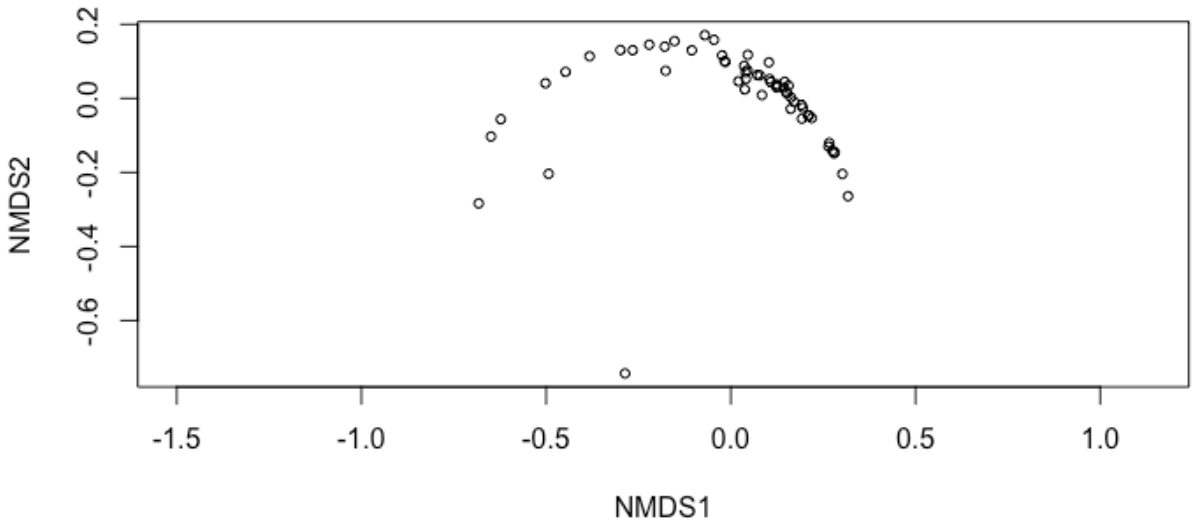


Figure 16: Pathogen Match NMDS Plot

Table 13: Abridged Relative Abundance Table- Qiime 2 Results

Kindgom	Phylum	Class	Order	Family	Genus	Species	HV0205	HV0219	HV0307	...	Average Relative Abundance
Fungi	Ascomycota	Dothideomycetes	Capnodiales	Cladosporiaceae	Cladosporium	unidentified	8.5%	9.5%	11.6%	...	10.9%
Fungi	Basidiomycota	Agaricomycetes	Russulales	Stereaceae	Stereum	unidentified	4.6%	10.5%	0.8%	...	6.0%
Fungi	Ascomycota	Dothideomycetes	Pleosporales	Didymellaceae	unidentified	unidentified	9.8%	26.7%	18.2%	...	5.8%
Fungi	Ascomycota	Dothideomycetes	Pleosporales	Pleosporaceae	Alternaria	Alternaria_eichhorniae	2.6%	1.0%	5.1%	...	3.5%
Fungi	Basidiomycota	Agaricomycetes	Polyporales	Coriolaceae	Trametes	Trametes_versicolor	18.6%	6.1%	2.3%	...	2.6%
Fungi	unidentified	unidentified	unidentified	unidentified	unidentified	unidentified	1.1%	0.5%	1.6%	...	2.5%
Fungi	Ascomycota	Sordariomycetes	Diaporthales	Gnomoniaceae	unidentified	unidentified	0.0%	0.0%	0.1%	...	1.8%
Fungi	Basidiomycota	Agaricomycetes	Agaricales	Tricholomataceae	Baeospora	Baeospora_myriadophylla	0.1%	0.0%	0.0%	...	1.6%
Fungi	Basidiomycota	Agaricomycetes	Auriculariales	Exidiaceae	Exidia	unidentified	6.3%	0.3%	0.2%	...	1.5%
Fungi	Ascomycota	Dothideomycetes	Pleosporales	unidentified	unidentified	unidentified	0.5%	0.3%	0.2%	...	1.4%
Fungi	Ascomycota	Dothideomycetes	Dothideales	Aureobasidiaceae	Aureobasidium	Aureobasidium_pullulans	1.3%	0.1%	1.4%	...	1.3%
Fungi	Ascomycota	Sordariomycetes	Xylariales	Diatrypaceae	unidentified	unidentified	0.0%	0.0%	0.0%	...	1.3%
Fungi	Ascomycota	Dothideomycetes	Pleosporales	Didymosphaeriaceae	unidentified	unidentified	0.1%	0.4%	0.6%	...	1.2%
Fungi	Basidiomycota	Agaricomycetes	Polyporales	Fomitopsidaceae	Sebipora	unidentified	0.1%	2.9%	0.3%	...	1.0%
Fungi	Ascomycota	unidentified	unidentified	unidentified	unidentified	unidentified	1.4%	0.2%	0.7%	...	1.0%
Fungi	Basidiomycota	Agaricomycetes	Hymenochaetales	Hymenochaetaceae	Hymenochaetopsis	Hymenochaetopsis_corrugata	0.4%	0.4%	0.1%	...	1.0%
Fungi	Basidiomycota	Ustilaginomycetes	Ustilaginales	Ustilaginaceae	Ustilago	Ustilago_hordei	0.0%	0.0%	0.0%	...	0.9%
Fungi	Basidiomycota	Agaricomycetes	unidentified	unidentified	unidentified	unidentified	0.3%	0.5%	0.6%	...	0.9%
Fungi	Basidiomycota	Agaricomycetes	Polyporales	Coriolaceae	Trametes	Trametes_cubensis	0.0%	0.0%	0.2%	...	0.9%
Fungi	Ascomycota	Dothideomycetes	Pleosporales	Periconiaceae	Periconia	Periconia_byssoides	0.2%	1.8%	0.3%	...	0.8%
Fungi	Basidiomycota	Agaricomycetes	Polyporales	Meripilaceae	Rigidoporus	Rigidoporus_pouzarii	1.0%	0.3%	2.9%	...	0.8%
Fungi	Ascomycota	Dothideomycetes	Capnodiales	Cladosporiaceae	Cladosporium	Cladosporium_sphaerospermum	2.8%	0.7%	2.8%	...	0.8%
Fungi	Basidiomycota	Exobasidiomycetes	Microstromatales	Microstromatales_fam	Pseudomicrostroma	Pseudomicrostroma_phylloplanu	0.2%	0.0%	0.1%	...	0.6%
Fungi	Ascomycota	Dothideomycetes	Capnodiales	Teratosphaeriaceae	Euteratosphaeria	Euteratosphaeria_verrucosiafrica	0.3%	0.2%	0.0%	...	0.6%
Fungi	Ascomycota	Dothideomycetes	Capnodiales	unidentified	unidentified	unidentified	0.4%	1.2%	1.1%	...	0.6%
Fungi	Basidiomycota	Agaricomycetes	Polyporales	unidentified	unidentified	unidentified	0.1%	0.9%	0.0%	...	0.6%